A Publication of the American Psychological Association

Journal of Abnormal and Social Psychology

Founded by Morton Prince, 1906

CONTENTS

Articles

The Process of Cognitive Tuning in Communication
Robert B. Zajonc

The Role of Uncertainty in Cognitive Change
Robert B. Zajonc and Julius Morrissette

Level of Stimulation and Memory Organization
Phillip S. Holzman and Riley W. Gardner

The Effect of Threat upon Interpersonal Bargaining
Morton Deutsch and Robert M. Krauss

Verbal Conditioning of Two Response Classes: Some Methodological Considerations
Joseph D. Matarazzo, George Saslow, and [illegible]

Anxiety in Academic Achievement Situations
Richard Alpert and Ralph Norman Haber

Effects of Waiting with Others on Changes in Level of Fear Anticipation
Lawrence S. Wrightsman, Jr.

The Influence of Some Types of Power Relationships and Game Strategies upon the Development of Interpersonal Trust
Leonard Solomon

Continued on inside front cover

Volume 61 SEPTEMBER 1960 Number

CONTENTS—*Continued*

Social Effectiveness and Symptomatic Behaviors 231
Edward Zigler and Leslie Phillips

Role Assignment and Interpersonal Influence 239
John D. Stoner and William L. Field

Vocabulary Knowledge and Usage of Schizophrenic Subjects: A Six-Year Follow-Up 246
Louis J. Moran, Donald R. Gorham, and Wayne W. Holtzman

A Microgenetic Approach to the Amnesic Syndrome 255
George A. Talland

The Adaptive Functions of Wit and Humor 263
Walter E. O'Connell

The Development of Meaning in Children's Descriptive Terms 271
Susan M. Ervin and Garrell Foster

A Therapeutic Approach to Schizophrenics of Extreme Pathology: An Operant-Interpersonal Method 276
Gerald F. King, Stewart G. Armitage, and John R. Tilton

Case Report

A Psychological Case Study of Death during Pregnancy 287
Anthony Davids and Morton Silverman

Critique and Notes

The Complementarity of Personality Needs in Friendship Choice 292
Mary St. Anne Reilly, William D. Commins, and Edward C. Stafford

Content Efforts in Social Perception 295
Leon H. Levy

The Formation of Positive Attitudes toward Group Members 297
Bernice Berman Lott and Albert J. Lott

Volume 61 September 1960 Number 2

Journal of Abnormal and Social Psychology

Editor: M. Brewster Smith, *University of California, Berkeley*
Associate Editor: E. J. Shoben, Jr., *Teachers College, Columbia University*

Consulting Editors

Gordon W. Allport
Harvard University

Urie Bronfenbrenner
Cornell University

Roger Brown
Massachusetts Institute of Technology

Donald T. Campbell
Northwestern University

Launor F. Carter
System Development Corporation

Richard Christie
Columbia University

Morton Deutsch
Bell Telephone Laboratories

I. E. Farber
State University of Iowa

Donald W. Fiske
University of Chicago

N. L. Gage
University of Illinois

Roger W. Heyns
University of Michigan

Robert R. Holt
New York University

Irving L. Janis
Yale University

Harold H. Kelley
University of Minnesota

Leonard S. Kogan
Community Service Society of New York

Richard S. Lazarus
University of California, Berkeley

Daniel J. Levinson
Harvard Medical School

David C. McClelland
Harvard University

Leo Postman
University of California, Berkeley

Eliot H. Rodnick
Duke University

Nevitt Sanford
University of California, Berkeley

Seymour Sarason
Yale University

Richard L. Solomon
University of Pennsylvania

PUBLISHED BY
THE AMERICAN PSYCHOLOGICAL ASSOCIATION, Inc.

The JOURNAL OF ABNORMAL AND SOCIAL PSYCHOLOGY is devoted to both abnormal and social psychology. Its emphasis is on basic research and theory rather than the techniques and arts of practice. Abnormal psychology is broadly defined to include papers contributing to fundamental knowledge of the pathology, dynamics, and development of personality or individual behavior, including deterioration with age and disease. Articles concerned with psychodiagnostic techniques are evaluated with respect to their contribution to an understanding of the psychological principles of diagnosis, those concerned with psychotherapy are judged in terms of their contribution to an understanding of the therapeutic process. Case reports are considered in terms of their heuristic value. Reports which serve to pose or to clarify important theoretical problems, or which promise to be useful in teaching, are often published. From the social area this JOURNAL gives preference to papers contributing to basic knowledge of interpersonal relations, and of group influences on the pathology, dynamics, and development of individual behavior.

The JOURNAL is issued bimonthly in January, March, May, July, September, and November, two volumes per year. The two volumes average 900 pages. The subscription price per year is $16.00; foreign, $16.50. Single issues are $3.00.

Original manuscripts should be sent to the Editor-elect:

DANIEL KATZ
Center for Advanced Study in
the Behavioral Sciences
202 Junipero Sera Boulevard
Stanford, California

ARTHUR C. HOFFMAN
Managing Editor

HELEN ORR
Promotion Manager

Subscriptions, address changes, and business communications should be sent to:

THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.
1333 Sixteenth Street, N.W.
Washington 6, D. C.

Changes of address must reach the Subscription Office by the 10th of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the post office that they will guarantee second-class forwarding postage. Other claims for undelivered copies must be made within four months of publication.

PUBLISHED BY
THE AMERICAN PSYCHOLOGICAL ASSOCIATION, Inc.
Mt. Royal and Guilford Aves., Baltimore 2, Md.
and 1333 Sixteenth Street N.W., Washington 6, D.C.

Second-class postage paid at Baltimore, Maryland, and at additional mailing offices

Copyright 1961 by the American Psychological Association, Inc.

Journal of Abnormal and Social Psychology
1960, Vol. 61, No. 2, 159–167

THE PROCESS OF COGNITIVE TUNING IN COMMUNICATION[1]

ROBERT B. ZAJONC
University of Michigan

CONSIDERABLE evidence is now available on selective and directive effects in perception and cognition. The earliest explanation of these effects was advanced by the Würzburg School in their postulation of a determining tendency that arises from the subject's (*S*'s) task assignment (*Aufgabe*), and acts as a steering mechanism in information-seeking behavior (Ach, 1905; Lewin, 1922a, 1922b; Watt, 1905). The specific effects of determining tendencies were shown to be intimately related to the nature of the *Aufgaben* (Chapman, 1932; Kulpe, 1904).

The effects of *Aufgaben* are readily detected in communication. Each person is daily a source and a destination of numerous messages, and strong selective and directive effects should result when the individual's role in the communication process is manipulated. Hovland and his associates (Hovland, Janis, & Kelley, 1953) have shown in extensive studies that the characteristics of the communicator and the circumstances of the communication influence the attitudinal and cognitive impact of communications. Zimmerman and Bauer (1956), on the other hand, have demonstrated that communicators themselves are influenced by their audiences. The purpose of this study is to examine the nature of cognitive structures that are activated or "tuned in" when persons enter into communication with others.

For the purposes of this research a set of formal properties were defined for the description and measurement of cognitive structures. Since it is generally agreed that cognitive structures represent organized systems whose nature depends on the various interrelations among their components (Asch, 1946; French, 1947), the main emphasis was placed on morphological description. The properties defined are to some extent based on Lewin's (1951) analysis of the life space and on French's (1947) work on sentiments. Two experiments were conducted in which measures of the cognitive dimensions were obtained and the effects of "tuning" to receive and transmit information were analyzed.

[1] Based upon dissertation in Social Psychology, University of Michigan, in 1954, in partial fulfillment of the requirements for the degree of Doctor of Philosophy. The author wishes to thank Dorwin Cartwright and Daniel Katz for their advice and suggestions in carrying out this study. This investigation was supported in part by a grant to the Research Center for Group Dynamics from the Rockefeller Foundation.

PROPERTIES OF COGNITIVE STRUCTURES[2]

It is assumed that objects and events are perceived and discriminated on the basis of psychological dimensions. By a psychological dimension is understood the organism's capacity to respond to stimuli in such a way that, given a set of stimuli and a set of responses made to them, the stimuli and responses form two ordered sets with a determinate correspondence between the elements of each set. From the person's responses one can infer the values of which psychological dimensions consist. An act of perceiving or discriminating may be thought of in terms of "projecting" the given stimulus on some psychological dimensions, and thereby attributing to it values taken from these dimensions. The values thus assigned to stimuli constitute the elements of cognitive structures and are called attributes. They are the qualities or characteristics that individuals "assign" to objects and events.

An attribute, then, is a concept (or a category) in its simplest form, since a value from a given psychological dimension generally characterizes an entire class of objects. More complex concepts are formed, as Bruner, Goodnow, and Austin (1956) suggest, by conjunctive and disjunctive combinations of attributes. The larger and the more inclusive the set of objects (class) denoted by a concept the more general is the concept. For instance, all sorts of objects are "black," "animal," or "beautiful." Only a few are "76 ft. 4$\frac{5}{32}$ ins. long."

A person has at his disposal a large number of concepts or attributes by means of which he can identify and discriminate objects and events in his environment. The set of all such attributes is called his cognitive universe. Let a cognitive structure then be defined as an organized subset of the given cognitive universe in terms of which the individual identifies and discriminates a particular object or event. The morphological properties of cognitive structures describe various relationships among attributes. Those defined here are differentiation, complexity, unity, and organization.

Degree of differentiation. Given a universe of let us say m alternative attributes by means of which a given

[2] For a more detailed exposition of the descriptive model see Zajonc (1954).

person is capable of identifying and discriminating objects and events, he can construct a maximum of 2^n distinct cognitive structures. The number (n) of attributes constituting a given cognitive structure is taken to reflect its degree of differentiation.

Degree of complexity. The attributes constituting a given cognitive structure may come from a single class or category of discriminanda, or they may represent many categories. For instance, a painting may be perceived exclusively in terms of its objective qualities, such as size, subject matter, color, type of frame, etc. Or, it may be looked upon partly in terms of its objective qualities, partly in terms of its formal properties (period, style, symmetry, harmony, etc.), and partly in terms of its impact on the subjective experience of the viewer. Such attribute groupings may be further subdivided into smaller classes, and the extent of such subdivision may be used in defining the complexity of cognitive structures.

Let us denote the level of inclusion of a given class by r, such that if the class K_i does not include another subclass, $r = 1$; when K_i includes some subclasses, which in turn do not include other subdivisions, $r = 2$; when K_i includes subclasses which themselves contain other indivisible subclasses, $r = 3$, etc. The degree of complexity of a given cognitive structure may be obtained by weighting each attribute by the level of inclusion of its class membership and summing the weighted scores thus obtained. Thus:

$$\text{com} = \sum_{r=1}^{n} r n_r$$

where n_r is the number of attributes in the rth level of inclusion.

Degree of unity. The structural components of a whole may depend on each other to a greater or lesser extent. The more attributes depend on each other the more the cognitive structure is unified. If we define the dependence of the attribute A_i on the attribute A_j as equal to 1 when a change in A_j produces a change in A_i, and as equal to 0 when change in A_j does not produce a change in A_i, then a dependency matrix can be constructed for all attributes of a given cognitive structure, and the total dependency of each attribute may easily be obtained by summing the entries in the appropriate row. To compare the unity of structures of different degrees of differentiation the measure of unity must be normalized. Given a structure with n attributes the maximum dependency of a given attribute, $\text{dep}(A_i)^{\max}$, is equal to $n - 1$, and the maximum sum of dependencies in the cognitive structure to $n(n - 1)$. Unity then can be defined as:

$$\text{uni} = \frac{\sum_{i=1}^{n} \text{dep}\ (A_i)}{n(n-1)}$$

where $\text{dep}(A_i)$ is the total dependency of the ith attribute.

Degree of organisation. Lewin (1951) thought of degree of organization of a whole in terms of the existence of a guiding force or principle that controls its parts. To the extent that one part or a cluster of parts dominates the whole, the whole is said to be highly organized. On the basis of this view the degree of organization can be readily defined. From the dependency matrix the degree to which a given attribute, A_i, depends on others can be computed, as well as the degree to which it determines others. This value is obtained by summing the entries in the appropriate column of the matrix. Taking the determinance of the strongest attribute, $\det(A_i)^{\max}$, and dividing it by unity of the cognitive structure we obtain a measure that reflects the degree to which the interdependence among the attributes is concentrated around a single core. Thus:

$$\text{org} = \frac{\det\ (A_i)^{\max}}{\text{uni}}$$

Empirical measures of the dimensions defined above can be obtained with the aid of various well-known methods that analyze the S's perception or conception of a given stimulus in terms of its characteristics or properties: adjective check lists, rating scales, the semantic differential, Q sort, and the like. The method used in the present series of studies required the S to describe a stimulus object, namely, another person, by freely listing the qualities and attributes that characterized him. The number of such qualities enumerated constituted the measure of differentiation. Complexity was measured by having the S organize the attributes he listed into meaningful groupings and subgroupings, which served as a basis for assigning weights reflecting the respective levels of inclusion. Both unity and organization were computed from Ss' ratings of dependencies among attributes they listed. Unity was scored by summing the dependency ratings among pairs of attributes, and, as required by the definition, dividing the sum by the maximum possible sum of dependencies. Organization was measured by finding for the given cognitive structure the attribute on which the greatest number of other attributes depended, taking its determinance score (i.e., simply the number of attributes that were rated as depending on it) and dividing it by unity.

The scope of this paper does not allow the discussion of the interrelations among the cognitive dimensions, either as these arise from the formal properties of the definitions or as they were observed in empirical results. An extended analysis of this problem is given elsewhere (Zajonc, 1954).

EXPERIMENT I

Although a person is capable of activating a large number of cognitive structures, at any one time only a few can operate. Specific conditions must therefore exist under which cognitive structures become activated, and the specification of these conditions should enable one to predict the nature of the structures that are activated. It is assumed that a cognitive structure relevant to a particular object or event is activated primarily when the person expects to deal with information about that object or event. Such anticipation constitutes one of the most fundamental occasions for the activation of cognitive structures. Casual observation reveals that messages exchanged between communicators at the onset of communication generally indicate what the communication is going to be about, and thus allow the activation of appropriate cognitive structures. One has merely to scan the opening paragraphs of articles in this journal to observe the tendency. The structure of language also shows such a propensity. Sentences generally begin with the subject. Placing the subject at the beginning of the sentence indicates what cognitive structures should be activated so that the information contained in the predicate can be meaningfully received.

If it is assumed that cognitive structures are activated in the anticipation of dealing with information, it follows that not only the subject matter of anticipated communication but also the person's role in the communication process determines the kind of cognitive structure that he activates. Two basic ways of dealing with information may be distinguished, receiving and transmitting. When a person primarily anticipates receiving information, he may be expected to activate a cognitive structure capable of admitting the incoming information. Concomitant with the anticipation of receiving information is the anticipation of cognitive change. On the other hand, anticipation of transmitting information should activate structures that may serve as a source of potential messages. These two processes may be termed receiving and transmitting tuning. The purpose of this study is to examine the differences in the properties of cognitive structures activated under receiving and transmitting tuning.

Method

Subjects. The *S*s used in this study were recruited on a voluntary basis from the AFROTC at the University of Michigan. Altogether 45 *S*s participated in the experiment.

Procedure. After assembling in a classroom, the *S*s in all conditions were issued the following verbal instructions by *E*:

> This is a study on how groups operate under certain conditions. I will tell you more about it later. Before, however, I will distribute copies of a real letter which was written by one individual to another. I want you to skim over this letter and get a general idea of what sort of person the writer is. Just try to imagine what kind of individual he is, and what are some of the things which are characteristic of him. But, please do not try to memorize the letter. This is not an experiment on memory. Just try to get a general picture of the individual who wrote the letter. You will have two minutes to read it.

At this point the letter was distributed face down; then a signal was given to begin reading the letter. After 2 minutes, a signal was given to turn the letter face down again. The copies of the letter were then collected.

The letter used in the experiment was written to a potential employer by a job applicant. It contained some of the qualifications of the applicant and information regarding his background.

Induction of cognitive tuning. Twenty *S*s received induction intended to produce transmitting tuning, and 25 received induction intended to produce receiving tuning. After the copies of the letter were collected the transmitters received the following instructions:

> I said that we are trying to discover how groups operate under certain conditions. We are especially interested in the process of communication.
>
> At this very moment there is another group in the building. Your responsibility will be to communicate the information you have obtained about the person who wrote the letter to the members of the other group. You will have to describe this person to the other groups, so that they can know him as well as you do now.
>
> But before we begin to transmit this information to the other group, we will put down the things we learned about the writer from his letter. I will now distribute forms which you will fill out according to the instructions written on each page.

The receivers were issued the same instructions except the second and third paragraphs were changed to read:

> At this very moment there is another group in the building. They have detailed information on the individual who wrote the letter, and they will communicate this information to you. We want to see how well they can convey to you all the information they have.
>
> But before we begin to receive this information from the other group, we will put down the things we learned about the writer from his letter. I will now distribute forms which you will fill out according to the instructions written on each page.

TABLE 1

DIFFERENCES IN THE PROPERTIES OF COGNITIVE STRUCTURES BETWEEN TRANSMITTERS AND RECEIVERS

Property	Mean Score for Transmitters	Mean Score for Receivers	t
Differentiation	7.90	5.12	5.67***
Complexity	17.80	9.32	6.83***
Unity	0.309	0.222	2.19*
Organization	13.75	8.40	2.94**

* $p < .05$.
** $p < .01$.
*** $p < .001$.

Measurement of the properties of cognitive structures. To obtain measures of the properties of cognitive structures the following procedure was used. *Ss* were issued instruction booklets to which were attached stacks of 52 blank cards marked from A to ZZ. The *Ss'* first task was outlined on the first page of the booklets:

Detach the stack of cards from this booklet.

On each card *separately* write *one* characteristic which describes the applicant. You can put down whatever comes to your mind, since there is no one list of characteristics that can be considered as either "correct" or "incorrect." Everyone of us sees things in a slightly different way.

You may have too many or too few cards, but this shouldn't bother you. Put down as many characteristics as you feel are necessary to describe the applicant adequately. Work rapidly.

The number of characteristics attributed to the applicant constituted the degree of differentiation in each case.

To determine complexity the *Ss* were required to:

Lay out in front of you all the cards you used for listing the characteristics of the applicant. Look over them carefully and notice whether they fall into some broad natural groupings. If they do, arrange them into such groups.

After arranging the cards into groupings, the *Ss* were issued the following instructions:

Now, look at your groups one by one and see whether these can't be broken down into subgroups. If they can, separate the cards accordingly. It is also possible that these subgroups can be broken down further, and so on.

When you arrange *all* cards into groups and subgroups, list your groupings on the sheet below as if they were points and subpoints of an outline. First, give names or titles to your groups and subgroups. Then in the right-hand column list the letters of all the characteristics that belong into the respective groups and subgroups.

The groupings and subgroupings were used to determine the levels of inclusion of the attributes, the basis for computing complexity scores.

The *Ss* were instructed to proceed as follows for the last task:

It is possible that some of the characteristics are related to one another. They may depend on one another in such a way that if one changes, the other ones would change too. Suppose the table in front of you were bigger than it is now. Then it would also become heavier. This means that the weight of the table depends on its size. The relationships between the characteristics you put down may not be so obvious and so simple, but try to decide whether such relationships exist nevertheless. To do this lay out the cards in front of you in alphabetical order, and follow the procedure below:

List all the characteristics which would change if Characteristic *A* were changed, absent, or untrue of the applicant; list all the characteristics which would change if Characteristic *B* were changed, absent, or untrue of the applicant, etc.

On the basis of the above responses dependency matrices were constructed and measures of unity and organization computed.

After the *Ss* completed the booklets they were told with an apology that the experiment was over.

Results and Discussion

Degree of differentiation. The overall results are shown in Table 1, which shows a significantly higher degree of differentiation for transmitters. The transmitters listed, on the average, 7.90 attributes, the receivers 5.12. It was assumed that the components of cognitive structures activated under tuning for reception act as *receiving categories*, whereas those activated under tuning for transmission act as *potential units of transmittable information*. The relevant difference between these two types of attributes is along the dimension of specificity–generality. Attributes whose function is to admit all possible types of information must be sufficiently broad and general. Since they are to serve as "codes" or "files" in terms of which ever new units of information are admitted, they must be sufficiently comprehensive so that all possible information can be appropriately categorized.

When the components of the cognitive structure are to serve as potential units of transmittable information they are more likely to be specific. A "message" is composed of specific units alone, whereas general units constitute skeletons that organize the substantive and specific units of the message. Since specific units require more detail, their essential difference from general units lies in the coarseness of their "grain." Cognitive representations made in specific terms normally require a larger number of attributes than representations made primarily in broad and general

terms. If it is assumed that, other things equal, differentiation is a function of the level of specificity of a given cognitive structure, and if tuning for transmission results in a tendency toward greater specificity than tuning for reception, transmitters should indeed show higher differentiation. Thus, if it can be demonstrated that the transmitters in fact employ a higher proportion of specific attributes than the receivers, then the differences in the degree of differentiation are better understood.

The attributes of all *S*s were coded for specificity by two independent judges, assigning each attribute to either specific or general categories. The general category included such items as "intelligent," "shiftless," "immature," "he does not seem to be ambitious." The specific category included attributes like, "didn't finish college because of financial problems," "he has never worked as a mail clerk." The reliability measured in percent agreement was 82%. The results indicate that the transmitters' cognitive structures had on the average 57.6% specific attributes, while those of the receivers, only 32.6%. This difference was significant at the .001 level.

Degree of complexity. The results show a higher degree of complexity for transmitters. Two factors may account for this finding. First, it would appear that the larger proportion of specific attributes, the more likely should the attributes fall into equivalence classes and subclasses. Furthermore, the results have indicated a higher degree of differentiation on the part of the transmitters. From the definition of complexity it is evident that a high degree of differentiation increases the probability of high complexity. Both factors work to enhance complexity on the part of the transmitters.

Degree of unity. From the concept of cognitive tuning it would follow that receiving tuning corresponds to the anticipation of a cognitive change in one's own cognitive structure, whereas transmitting tuning corresponds to the anticipation of inducing changes in cognitive structures of other persons. A person who anticipates the occurrence of cognitive change is very likely to "unfreeze" his cognitive structure or to make it more susceptible to the anticipated changes. Susceptibility and resistance to change may be thought of in terms of the strength of forces necessary to induce changes in a given cognitive structure. In order to change a highly unified structure a strong force is necessary because the components of the structure depend on one another to a great degree. Thus a change in one attribute should result in changes in other attributes. In order to change a single attribute of a highly unified cognitive structure, a force must be applied which is capable of overcoming the resistance of not only the given attribute, but of all the other attributes on which it depends. To produce a similar change in an un-unified structure, a force is required that is capable of overcoming the resistance of the given component alone. If cognitive structures activated under receiving tuning are more susceptible to change than those activated under transmitting tuning, then they must be less well unified than the latter. The results indicate that this, in fact, is the case.

Degree of organization. The dynamics of receiving tuning are of such nature that a high degree of structure and elaboration should deter the smooth reception of information. Rather, the situation calls for unstructured and loosely integrated cognitive structures, so that all possible types of changes could easily be introduced. The receiver does not know what specific information is to be imparted to him, and he cannot focus his cognitive processes on any single aspect of the matter, but must think in general terms. Degree of organization, it will be recalled, was defined in terms of the extent to which the cognitive structure is organized around such one specific aspect. Receiving tuning should therefore result in a relatively low degree of organization.

Transmitting tuning, however, represents a different psychological state. The transmitter is aware of the specific units of information to be transmitted, and his thinking revolves around them, with a certain focus of attention. He should therefore show a relatively higher degree of organization.

One should also expect a higher degree of organization for transmitters on other grounds. Predominance of specific attributes requires a structure into which the attributes can be ordered. The specific attributes are organized by means of general ones. Since the transmitters were shown to employ fewer general

attributes, their cognitive structure should be organized around fewer foci than those of the receivers.

The results show that the degree of organization was significantly higher for transmitters than for receivers. It should be noted that this result was obtained independently of the differences in unity. From the definition it follows that, other things being equal, degree of organization actually decreases with higher unity. The higher degree of organization found for the transmitters was obtained despite their higher unity.

EXPERIMENT II

In the previous experiment lower degrees of differentiation, complexity, unity, and organization were obtained under receiving than under transmitting tuning. These results were assumed to derive from the fact that the receivers tune in flexible cognitive structures, readily susceptible to changes, and having few specific attributes, while the transmitters' cognitive structures are characterized by a considerable resistance to change and a larger proportion of specific attributes. If resistance to change and degree of specificity account for the differences arising under receiving and transmitting tuning, they should also account for data obtained under other conditions involving these two factors. Suppose that persons expect to enter into communication with others who hold opinions opposite to or incongruent with their own. In the case of receiving tuning, there should be more resistance to accepting information that is contrary to the person's own beliefs than to accepting information that supports them. Since resistance to cognitive change has been assumed to be reflected in increased unity, we expect the receivers to show a relatively higher degree of unity when dealing with incongruent information than when dealing with congruent information. Furthermore, if they are to protect themselves from unwanted changes, receivers of incongruent information should eliminate from their cognitive structures components that make them vulnerable to change. If they have committed themselves to a positive opinion, they should ignore or omit negative attributes, and if they have committed themselves to a negative opinion, the opposite should be the case. Jones and Kohler (1958) have recently shown that *S*s tend to recall plausible arguments that support their opinion and implausible arguments that question it, and to forget the converse. Simultaneously, the commitment to an opinion places a requirement on specificity. Thus the first factor, because of its *selective* effects should tend to reduce differentiation for receivers of incongruent information, while the second should tend to increase it. If the two factors are equal in magnitude, the degree of differentiation should remain the same for receivers regardless of whether they deal with congruent or incongruent information. Similar consequences should follow for transmitters. The differences in the degree of differentiation between receivers and transmitters when incongruent information is involved should depend therefore on the relative increments of specificity and selectivity in both groups. If we assume that both groups select all those and only those attributes which are consistent with their position, then because the number of such attributes is held constant for both groups, no differences in specificity should appear, and, consequently, the groups should not manifest any differences in differentiation. If receivers and transmitters increase in specificity and selectivity when confronted with incongruent information, with these increments being more pronounced for receivers, we should expect a lowering of differentiation for transmitters and a heightening for receivers. The differences in complexity should show the same pattern as differences in differentiation. Degree of organization, on the other hand, should increase for both groups, since now there is a core around which cognitive components may readily become organized—namely, the commitment to a positive or a negative position. But again this increase should be relatively more pronounced for the receivers, resulting in approximately equal levels of organization for both groups. The purpose of the second experiment is to examine these predictions.

Method

Subjects. The *S*s were sophomore and junior women enrolled at the University of Michigan. Altogether, 92 *S*s took part in the experiment.

Procedure. Four groups of 23 *S*s each were assigned

randomly to four experimental groups: Transmitters of Congruent Information (T), Receivers of Congruent Information (R), Transmitters of Incongruent Information (TI), and Receivers of Incongruent Information (RI). The stimulus materials and the measurement method were the same as in Experiment I. Also, the procedure for the T and R groups was the same as in Experiment I, except that for the present purposes the induction of tuning preceded the reading of the letter. The TI and RI groups were issued the following instructions after assembling in the experimental room:

> The main point of this study is to find out something about the process of communication. I will give you a letter to read. This letter was written by an individual who is applying for a job. After you have read this letter and have gotten some idea about its author, each one of you will decide for herself whether or not he should be hired. Let me tell you something else. Right now there is another group in the building. They have the same information about the applicant, and each of them is also making a decision whether or not the individual should be hired. After you have decided whether or not to hire the applicant, every one of you will be paired with a member of the other group. What is more, you will be paired up with someone who decided differently from you. In other words, if a member of this group had decided to hire him, she will be paired with someone who has decided not to hire him, and vice versa.

Both the TI and the RI groups were given these initial instructions, and, afterwards, transmitting and receiving tuning was induced in the same manner as it was in Experiment I. The procedure for collecting data was the same as in the previous experiment.

Results

It was assumed that anticipation of dealing with incongruent information and commitment to an opinion result in an increased proportion of specific attributes, especially for the receivers. The average percentages of specific attributes for the T, R, TI, and RI groups, respectively, were 52.0%, 39.8%, 72.3%, and 70.7%. The analysis of variance (using the arc sin $\sqrt{\%}$ transformation) presented in Table 2 shows significant effects for both tuning and congruence. It will be noted that under conditions that involve anticipation of dealing with incongruent information, both receivers and transmitters increase the proportion of specific attributes markedly, with the increase being somewhat larger for the receivers. It is also clear from the data that when persons anticipate dealing with contrary information, the differences in the proportion of specific attributes between the groups disappear. The difference in the proportion of specific components between the R and the T groups was significant at the .01 level, and not significant between the RI and TI groups.

Degree of differentiation. The increment in specificity alone would lead to the expectation that the degree of differentiation should be greater for both groups that dealt with incongruent information. The receivers of contrary information should show a more pronounced increase, such that they reach the level of differentiation equal to that of the TI group. However, it was assumed that a factor operates simultaneously that tends to reduce differentiation, namely, the tendency to select attributes consistent with one's position. If this factor operates equally for both groups, as was assumed above, the TI groups should show a decrement and the RI group an increment. The results for all the properties, as summarized in Table 3, follow just this pattern. The analysis of variance, presented in Table 4, shows significant main effects deriving from tuning as well as a significant T × C interaction. The overall mean differentiation scores are somewhat higher than when *S*s deal with congruent information, but this difference is not significant. The differences between transmitters and receivers are significant only when congruent information is involved ($p <$

TABLE 2

SUMMARY OF ANALYSIS OF VARIANCE FOR THE PROPORTION OF SPECIFIC ATTRIBUTES

Source	*df*	*MS*	*F*
Tuning (T)	1	3,885.61	5.85*
Congruence (C)	1	8,068.15	12.14***
T × C	1	1,585.74	2.39
Error	88	664.27	

* $p < .05$.
*** $p < .001$.

TABLE 3

DIFFERENCES IN THE PROPERTIES OF COGNITIVE STRUCTURES BETWEEN TRANSMITTERS AND RECEIVERS DEALING WITH CONGRUENT AND INCONGRUENT INFORMATION

Property	Transmitters		Receivers	
	Congruent	Incongruent	Congruent	Incongruent
Differentiation	11.39	9.26	6.65	7.70
Complexity	27.09	20.83	14.35	16.48
Unity	0.265	0.269	0.226	0.300
Organization	15.54	22.13	10.41	19.48

TABLE 4
SUMMARY OF ANALYSES OF VARIANCE FOR DATA IN TABLE 3

Source	df	Differentiation		Complexity		Unity		Organization	
		MS	*F*	*MS*	*F*	*MS*	*F*	*MS*	*F*
Tuning (T)	1	228.53	30.92***	1678.79	21.15***	.027	1.35	568.13	7.12**
Congruence (C)	1	6.79	<1.00	198.10	2.50	.035	1.75	1034.51	12.97***
T × C	1	64.72	9.13***	304.88	3.48*	.001	<1.00	151.33	1.90
Error	88	7.39		79.36		.020		79.75	

* $p < .05$ level.
** $p < .01$ level.
*** $p < .001$ level.

.001). It appears, therefore, that there must exist strong selective effects, as well as increments in specificity when persons deal with contrary information, and that these two factors result in the equalization of differentiation.

Degree of complexity. The results for complexity are essentially the same as for degree of differentiation. The effects of tuning on complexity were significant ($p < .001$), as was the T × C interaction ($p < .05$). The degree of complexity of the transmitters shows a drop when they anticipate dealing with contrary information. The difference between the T and the TI group is significant at the .01 level. The decrement in complexity on the part of the transmitters is due to the selective effects of commitment to a position. Since they abandon those aspects that may weaken their position, their cognitive structures lose in elaboration. The receivers show a slight but insignificant increment in complexity, and a significant T × C interaction is thus obtained. It should be noted, however, that the TI group shows a higher degree of complexity than the RI group, this difference being significant at the .05 level. Apparently, the effects of tuning are sufficiently strong here to be maintained despite the selectivity and the specificity factors.

Degree of unity. While there is some increase in the degree of unity in the "incongruent" condition, the overall effects of congruence do not reach an acceptable level of significance. The strongest difference appears to be that between the R and the RI group. This was the only significant result with respect to unity ($p < .05$). It may be assumed, therefore, that when dealing with incongruent information receivers protect themselves from unwanted changes by increasing unity, while transmitters because of their already high unity need not increase it.

Degree of organization. As expected, degree of organization shows pronounced effects of both tuning and congruence. For both, receivers and transmitters, there are significant increments in organization when contrary information is involved (both differences are significant at the .01 level). The commitment to a position, therefore, because it leads to the selection of a particular type of material—selection that is quite systematic—provides a strong core around which the components of the cognitive structure may become readily organized. Even though the receivers showed an increase in unity when dealing with contrary information, they manifest a significant, almost twofold, increment in degree of organization. The effects of the anticipation of dealing with incongruent information are therefore stronger for the receivers than for transmitters.

SUMMARY

The cognitive effects of the person's role in the communication process were examined in two experiments. A systematic method for the description of cognitive structures was developed, and in the first experiment persons expecting to transmit information were compared with others expecting to receive information for the extent of differentiation, complexity, unity, and organization. The results show that transmitters activate cognitive structures which are more differentiated, complex, unified, and organized than those activated by receivers.

The second experiment involved the antici-

pation of dealing with incongruent information, and in comparison with groups dealing with congruent information, those expecting to deal with incongruent information generally showed decreased differences between transmitters and receivers. These results were accounted for by an increase in the proportion of specific cognitive components and in the tendency to reject material inconsistent with the person's own opinion.

REFERENCES

ACH, N. *Ueber die Willenstätigkeit und das Denken.* Gottingen: Yandenhoeck & Ruprecht, 1905.

ASCH, S. E. Forming impressions of the personality. *J. abnorm. soc. Psychol.*, 1946, **41**, 258–290.

BRUNER, J. S., GOODNOW, J., & AUSTIN, G. *A study of thinking.* New York: Wiley, 1956.

CHAPMAN, D. W. Relative effects of determinate and indeterminate Aufgaben. *Amer. J. Psychol.*, 1932, **44**, 163–174.

FRENCH, V. The structure of sentiments. *J. Pers.*, 1947, **15**, 247–280.

HOVLAND, C. I., JANIS, I. L., & KELLEY, H. H. *Communication and persuasion: Psychological studies of opinion change.* New Haven: Yale Univer. Press, 1953.

JONES, E. E., & KOHLER, R. The effects of plausibility on the learning of controversial statements. *J. abnorm. soc. Psychol.*, 1958, **57**, 315–320.

KULPE, O. Versüche über Abstraction. *Ber. I. Kongr. Exp. Psychol.*, 1904, 56–68.

LEWIN, K. Das Problem der Willenmessung und das Grundgesetz der Assoziation. *Psychol. Forsch.*, 1922, **1**, 191–302. (a)

LEWIN, K. Das Problem der Willenmessung und das Grundgesetz der Assoziation. *Psychol. Forsch.*, 1922, **2**, 65–140. (b)

LEWIN, K. *Field theory in social science.* New York: Harper, 1951.

WATT, H. J. Experimentelle Beiträge zur einer Theorie des Denkens. *Arch. ges. Psychol.*, 1905, **4**, 289–436.

ZAJONC, R. B. Cognitive structure and cognitive tuning. Unpublished doctoral dissertation, University of Michigan, 1954.

ZIMMERMAN, CLAIRE, & BAUER, R. A. The effect of an audience upon what is remembered. *Publ. opin. Quart.*, 1956, **20**, 238–248.

(Received August 10, 1959)

Journal of Abnormal and Social Psychology
1960, Vol. 61, No. 2, 168-175

THE ROLE OF UNCERTAINTY IN COGNITIVE CHANGE[1]

ROBERT B. ZAJONC AND JULIAN MORRISSETTE[2]

University of Michigan

In analyzing the process of social change in small groups Lewin (1952) proposed that "a successful change includes . . . three aspects: unfreezing (if necessary) the present level L_1, moving to the new level L_2, and freezing group life on the new level." He argued that since stable states resist changes the stability of the group must be reduced before a permanent change can be introduced. These components of the process of changing appear to have a rather general application. Individual as well as social change might be so characterized. The processes of attitudinal and cognitive change, for instance, may well involve the three steps distinguished by Lewin. The literature on the Water-Jar Einstellung Test (Levitt, 1956) shows some evidence that the ability to see the new solution is related to variability on previous solutions. That is, the likelihood of a new strategy depends on the rigidity of the old strategy.

The purpose of this study is to determine whether cognitive change involves similar dynamics. The specific cognitive variable to be examined is cognitive uncertainty. It has been shown that cognitive change and cognitive uncertainty are intimately related, and in particular, that the likelihood of cognitive change increases with uncertainty (Zajonc & Burnstein, 1957). It is not clear, however, whether cognitive change occurs *only* with a concomitant increase in uncertainty. According to Lewin's conception of the dynamics of change, it is hypothesized that uncertainty must increase before a cognitive change can take place. Of course, when the person is very uncertain to begin with, no such increase is to be expected, and we should therefore expect that increases in uncertainty accompanying cognitive change are more likely to occur at low levels of uncertainty. But given low uncertainty, cognitive change can be accomplished only by previously raising uncertainty.

[1] These studies were a part of a larger program supported by Grant G-2802 from the National Science Foundation.

[2] Now at Wright-Patterson Air Force Base.

METHOD

Subjects

Ninety-six male *Ss* recruited on a volunteer basis from the undergraduate classes at the University of Michigan participated in the experiment. Eight groups of 12 *Ss* each were employed.

Procedure

In all groups the *Ss*' task was to estimate the number of bomb craters shown in a slide of an aerial photograph. In each group, the slide was presented 18 times tachistoscopically for intervals of 0.5 sec.

Three methods of inducing cognitive change were employed: No Feedback (NF), Feedback I (FI), and Feedback II (FII). In the NF condition *Ss* estimated the numbers of bomb craters on each of the 18 presentations and no feedback as to their accuracy was ever given. Three experimental groups constituted the NF condition, NF_a, NF_b, and NF_c. In the NF_a group the same slide was presented in all of the 18 trials. In the NF_b group the experimenter (*E*), without the *Ss*' awareness, changed the slide following the sixth trial, and in the NF_c following the twelfth trial. The original slide pictured 350 craters, the new slide 250. Both slides showed the same area, but in the second slide 100 bomb craters were eliminated by retouching.

In the FI groups the slide containing 350 craters was presented throughout all 18 trials. After each trial the *E* would announce a number representing the mean estimate of a group of pilots who supposedly were previous *Ss* of the experiment. Again three experimental groups were employed: FI_a, FI_b, and FI_c. In the FI_a group the estimate reported by the *E* was always around 350. The mean of the estimates given by the *E* over the 18 trials was 350. On any trial the deviation from 350 was never greater than 10. In the FI_b group the reported mean changed from 350 to 250 following the sixth trial, and in the FI_c following the twelfth trial.

In the FII condition the feedback was held constant at around 350 craters for all 18 trials. In these groups, however, the slide was changed as in the NF condition. Since the FI_a group provides a baseline for the FII groups, only two experimental groups were employed, FII_b and FII_c. In the FII_b the slide changed following the sixth trial, and in the FII_c following the twelfth trial.

In all groups the *Ss* were required to write down their judgments as to the number of bomb craters presented and also to indicate the smallest range within which their estimates would be perfectly accurate. This procedure provided a subjective measure of uncertainty (SU) which reflects the *Ss*' own estimates of their errors. Another measure of uncertainty was obtained by employing the variability of *Ss*' estimates as to the number of bomb craters. The trials were grouped

into six trial blocks consisting of three successive trials each, and for each *S*, average deviations of the three estimates within each trial block were computed. This latter, more objective, measure of uncertainty will be referred to as the OU measure.

RESULTS AND DISCUSSION

Decrement in Uncertainty (OU Measure) as a Function of Repeated Trials

Before considering the role of uncertainty in cognitive change the data from the two control groups, NF_a and FI_a will be examined. The results for the two groups are presented in Figure 1 and the analysis of the data in Table 1.

In both F and NF conditions, there was a pronounced decrement in uncertainty as a function of repeated exposures of the stimulus. The analysis of variance shows an *F* ratio of 15.00, significant at the .001 level, for differences between Trial Blocks. It is also clear from Figure 1 that while the FI_a group shows a considerably higher level of uncertainty at the outset of the session, its uncertainty falls below that of the NF_a group beginning with the second trial block. The *t* test between the two groups on Trial Block I was significant at the .01 level. Thus we should expect a significant interaction between the effects of Feedback and Trial Blocks, as the analysis of variance in fact confirms.

TABLE 1
ANALYSIS OF VARIANCE FOR DATA IN FIGURE 1

Source	*df*	*MS*	*F*
Feedback	1	881	<1.00
Ss in same group	22	1,002	
Trial blocks	5	12,010	15.00***
Trial blocks × Feedback	5	2,298	2.87*
Pooled *Ss* × Trial blocks	110	800	

* $p < .05$.
*** $p < .001$.

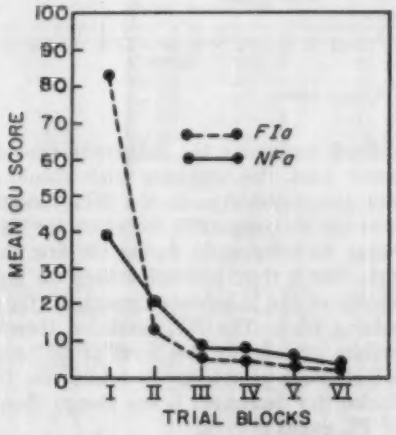

FIG. 1. Changes in uncertainty (OU Measure) as a function of repeated exposures of the stimulus.

TABLE 2
ANALYSIS OF VARIANCE FOR DATA IN FIGURE 2

Source	*df*	*MS*	*F*
Feedback	1	158	<1.00
Ss in same group	22	52,234	
Trial blocks	5	20,778	7.83***
Trial blocks × Feedback	5	8,909	3.36**
Pooled *Ss* × Trial blocks	110	2,654	

** $p < .01$.
*** $p < .001$.

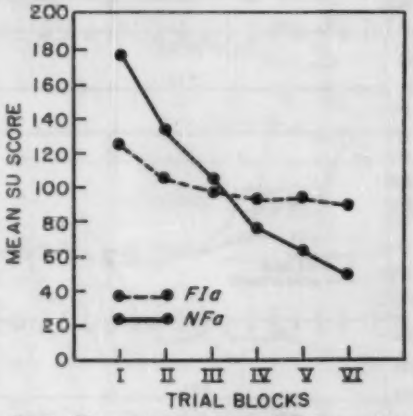

FIG. 2. Changes in uncertainty (SU Measure) as a function of repeated exposures of the stimulus.

Decrement in Uncertainty (SU Measure) as a Function of Repeated Trials

When uncertainty is measured by the *Ss*' estimates of their error, the decrement over Trial Blocks is again observed, but now the interaction between Feedback and Trial Blocks is reversed. It is clear from Figure 2 that the FI_a curve has a lesser slope than the NF_a curve, and that at the outset uncertainty of the NF_a group is higher than that of the FI_a group. The analysis of variance presented in Table 2 indicates again a significant effect deriving from Trial Blocks as well as a significant interaction between Trial Blocks and Feedback.

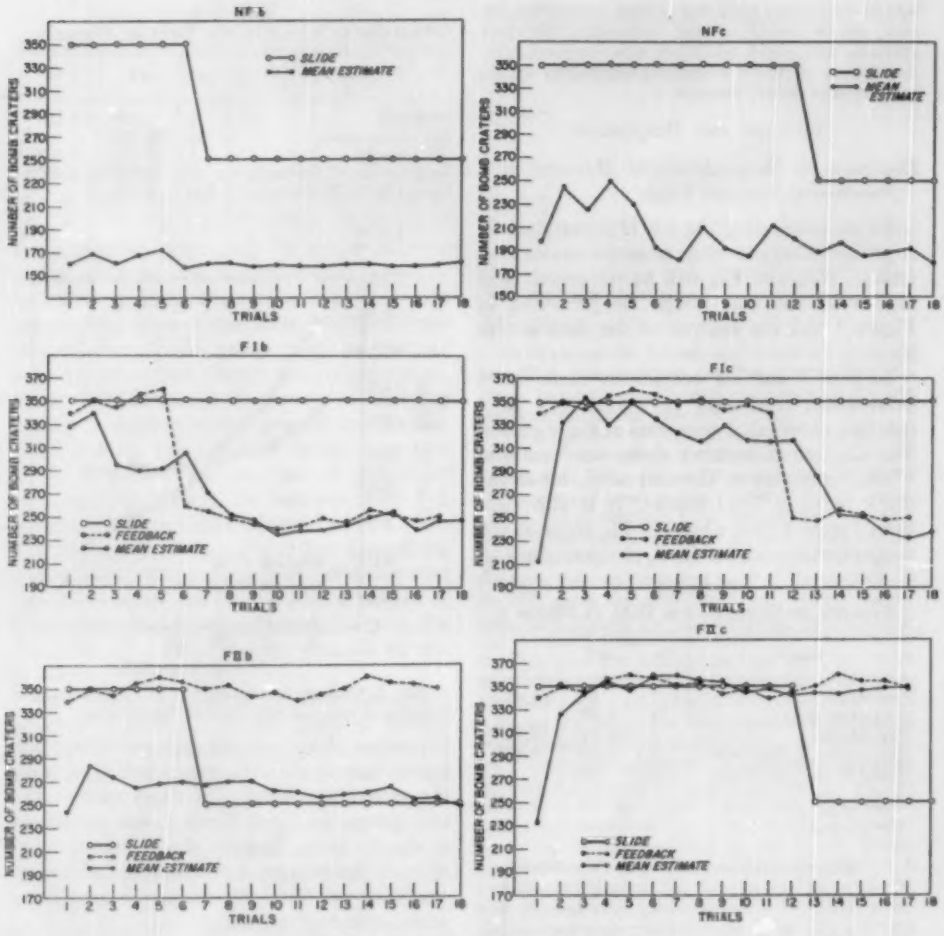

FIG. 3. Mean estimates of bomb craters.

The difference in the direction of interaction is probably a function of the uncertainty measures employed. Since the OU measure is essentially a measure of intertrial and intrasubject variability we should expect that feedback would originally produce high judgmental variability and thus high OU scores. The first feedback occurs after the first trial of the first trial block, after *S* has already made his judgment. If feedback has any effect on *S*'s judgments it should result in a change of his judgment on the immediately subsequent trials, and therefore a high average deviation on the first trial block. Moreover, since the feedback maintains the judgments around a certain level, the remaining trials should exhibit low variability. In the NF_a condition there are no comparable forces on the *S*s to change their estimates during the first three trials. Nor is there pressure toward the maintenance of the judgments throughout the remaining trials. The NF_a condition therefore exhibits an initially low level of OU scores and, although uncertainty decreases over Trial Blocks, this decrement is less abrupt than in the FI_a group.

The SU measure, however, reflects the *S*s' judgmental confidence. If feedback serves as a

source of additional information it should increase confidence and result in the FI_a group in relatively low SU scores on the initial trials. Thus, during the first three trial blocks the SU scores are lower for the FI_a group than for the NF_a group. However, the FI_a group does not decrease as rapidly as the NF_a group, a result that is difficult to interpret. It implies that acceptance of the additional information provided by the feedback may reduce uncertainty at the outset, and simultaneously acts to maintain the resulting level of uncertainty through the remainder of the session.

Uncertainty and Cognitive Change

The data thus far presented show that repeated exposures to a stimulus result in uncertainty that decreases progressively. Both the person's judgmental variability and his subjective estimate of error decrease significantly over repeated trials. The "unfreezing" hypothesis suggests that if the individual faces a change of judgment, this change is concomitant with increased uncertainty. Moreover, if we assume that there exists a level of uncertainty at which cognitive change can readily occur, and below which cognitive change requires an increase in uncertainty, it follows

TABLE 3

MEAN ESTIMATES OF BOMB CRATERS BEFORE AND AFTER STIMULUS AND FEEDBACK CHANGE

Condition	Mean Estimate of Bomb Craters		Difference
	Before	After	
NF_b	171.42	159.67	−11.75
NF_c	209.67	184.08	−25.59
FI_b	308.33	247.83	−60.50
FI_c	321.92	251.17	−70.75
FII_b	271.08	259.92	−11.16
FII_c	339.42	346.17	6.75

TABLE 4

ANALYSIS OF VARIANCE FOR DATA IN TABLE 3

Source	*df*	*MS*	*F*
Feedback (F)	2	206,499	24.03***
Time of change (T)	1	4,834	<1.00
T × F	2	14,747	1.72
Ss in same group	66	8,594	
Trial blocks (B)	1	29,929	9.01**
B × F	2	12,995	3.91**
B × T	1	38	<1.00
B × F × T	2	909	<1.00
Pooled *Ss* × B	66	3,322	

** $p < .01$.
*** $p < .001$.

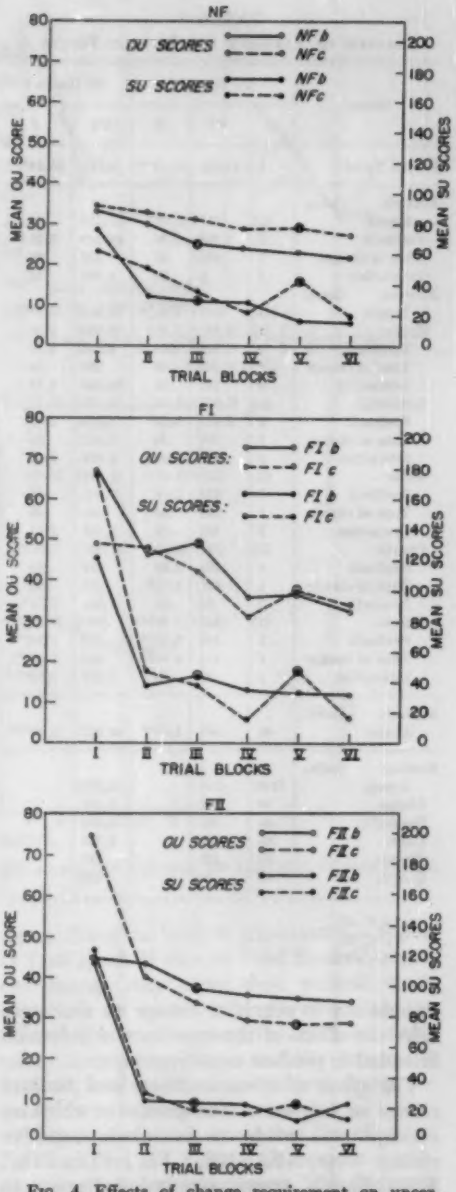

FIG. 4. Effects of change requirement on uncertainty. (Points circled indicate where the requirement for change was introduced.)

that a cognitive change on the third trial block should be accompanied by a lesser increase in uncertainty than on the fifth trial block. However, before introducing results on the role of

TABLE 5
ANALYSIS OF VARIANCE FOR DATA IN FIGURE 4

Source	OU Measure			SU Measure	
	df	*MS*	*F*	*MS*	*F*
Overall Trend	5	13,612	35.59***	34,874	16.03***
Between Group Means	(5)	(857)	(1.23)	21,717	.44
Feedback	2	1,930	2.76	49,315	1.01
Time of change	1	425	.61	137	.00
Interaction	2	1	.00	4,913	.10
Between Group Trends	(25)	(615)	(1.61)*	(3,406)	(1.57)*
Linear	(5)	(1,057)	(1.26)	(8,040)	1.42
Feedback	2	1,733	2.06	7,893	1.39
Time of change	1	798	.95	583	.10
Interaction	2	511	.61	11,916	2.10
Quadratic	(5)	(1,023)	(1.76)	(4,674)	(1.77)
Feedback	2	2,062	3.54*	4,774	1.81
Time of change	1	209	.36	1,652	.62
Interaction	2	392	.67	6,086	2.30
Cubic	(5)	(502)	(1.67)	(1,884)	(1.09)
Feedback	2	921	3.06	1,066	.62
Time of change	1	248	.83	165	.09
Interaction	2	209	.70	3,561	2.07
Quartic	(5)	(331)	(2.31)	(1,433)	(1.89)
Feedback	2	405	2.83	522	.69
Time of change	1	678	4.74*	419	.55
Interaction	2	83	.58	2,851	3.77*
Quintic	(5)	(162)	(3.49)**	(999)	(7.51)***
Feedback	2	464	9.98***	792	5.95**
Time of change	1	228	4.91**	936	7.03**
Interaction	2	27	.59	1,237	9.30***
Between Indiv. Means	66	699	1.83**	49,006	22.52***
Between Indiv. Trends	(330)	(383)		(2,176)	
Linear	66	841		5,667	
Quadratic	66	582		2,644	
Cubic	66	301		1,724	
Quartic	66	143		757	
Quintic	66	46		133	

* $p < .05$.
** $p < .01$.
*** $p < .001$.

uncertainty in cognitive change we shall consider the effects of the experimental induction intended to produce cognitive change.

The effects of stimulus change and feedback change on judgments. The groups in which an attempt was made to introduce cognitive change were NF_b, NF_c; FI_b, FI_c; FII_b, FII_c. The b groups received induction to change on the third trial block and the c groups on the fifth. In the NF condition the experimental induction of cognitive change was introduced by changing the slide. In the FI groups the slide remained the same throughout the session but feedback changed. And in the FII groups the slide changed while feedback remained constant. The *Ss*' average estimates of the number of bomb craters for the individual trials are presented in Figure 3. It is clear from the data that the only groups where change of any magnitude occurred as a result of the induction were the FI_b and FI_c groups. It is also evident that in all but the FII_b groups there was a general tendency to conform to the information provided in feedback rather than to that given in the slide.

The results pertaining to cognitive change were evaluated by comparing estimates for the two trial blocks in which change induction was introduced. The average estimates for the trial block immediately preceding and the trial block immediately succeeding the change induction are shown in Table 3 and the analysis of variance on these results in Table 4. The data of main interest are of course the mean estimates for the critical trial blocks. It is clear that there was some overall change in the direction of the slide and feedback changes. The differences between the critical trial blocks are significant at better than the .01 level. The strongest difference, however, occurred in the FI condition and the weakest in the FII condition. The interaction between Trial Blocks and Feedback is significant at the .05 level. These results suggest that the effects stemming from Feedback were more pronounced than those from the stimulus itself. Since the *Ss* viewed the slide under suboptimal conditions, they would, of course, tend to rely more on information provided in Feedback. The analysis also shows that in the absence of Feedback *Ss* generally underestimated the number of bomb craters in the aerial photograph, and consequently the overall effects of Feedback are significant. The difference in estimates between Trial Blocks IV and V is somewhat greater than between Trial Blocks II and III, but the interaction between Time of Change and Trial Blocks was not significant.

Changes in uncertainty as a function of stimulus change and feedback change. The results for judgments showed that only the FI groups underwent a cognitive change of any significance. The NF and the FII groups changed their estimates only slightly with the FII condition showing the weakest change. We would therefore expect that increments in uncer-

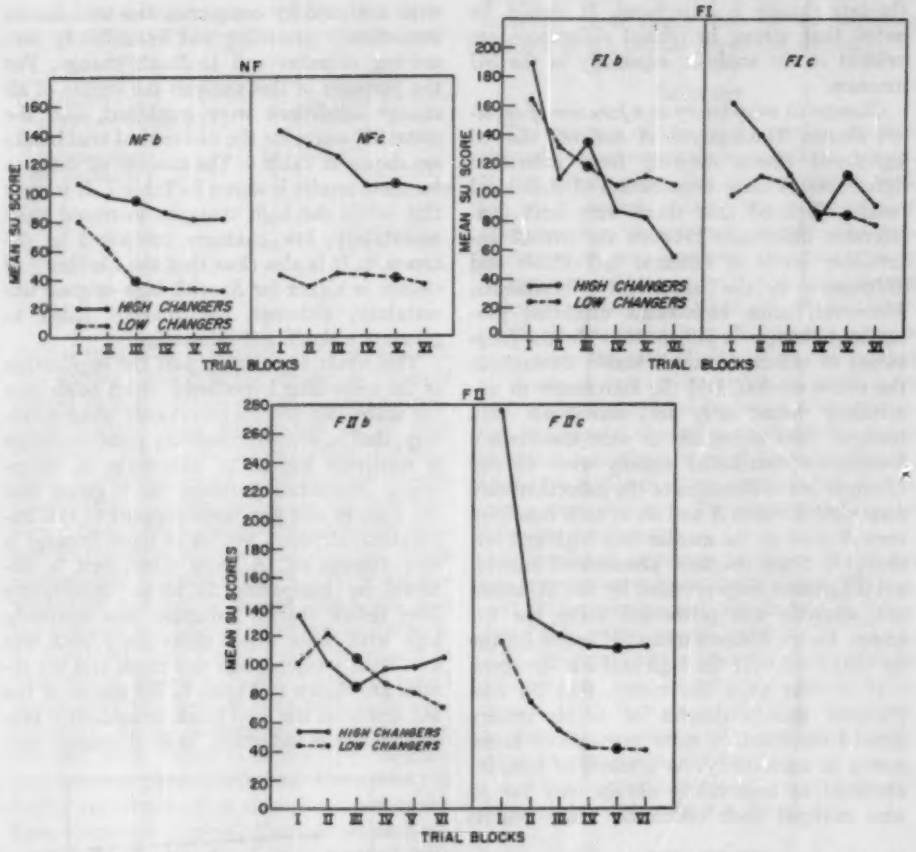

FIG. 5. Effects of change requirement on uncertainty (SU scores) for high and low changers. (Points circled indicate where the requirement for change was introduced.)

tainty at the point of change would occur in corresponding order. The data for all groups in which the attempt was made to introduce cognitive change are shown in Figure 4. Since the hypothesis predicts a departure from monotonicity, orthogonal polynomial trend analyses (Grant, 1956) were performed on the data and the results are shown in Table 5.

The findings are essentially the same for the OU and the SU scores. The curves show that when change in the slide or in feedback was introduced early during the session, a rise in uncertainty occurred only in the FI_b group. However, when the requirement for change occurred during the fifth trial block, all conditions showed an increment in the OU scores, and all but the FII_c group increments in SU scores. Since the level of uncertainty is higher at Trial Block II than at Trial Block V, cognitive change may occur then without much increment in uncertainty. However, at the close of the session when the level of uncertainty is at a minimum and the judgments have become stabilized, cognitive change requires an increase in uncertainty.

The trend analysis shows that the effects of Feedback and the Time of Change are significant for the quintic component. This finding is consistent with the prediction since the summated curves for the b and c groups generate a function with three maxima. One maximum obtains at Trial Block I, the other at Trial Block III where the early change is introduced, and the third at Trial Block V, where

the late change is introduced. It should be noted that strong individual differences are evident in the analysis, especially in the SU measure.

Changes in uncertainty as a function of cognitive change. The analysis of variance showed significant effects deriving from individual differences. A close examination of individual results disclosed that there were both considerable differences between the overall uncertainty levels of different individuals and differences in the uncertainty increments. Moreover, some individuals exhibited pronounced changes in judgments while the judgments of others remained stable throughout the entire session. Did the increments in uncertainty occur only for individuals who changed their estimates, or were uncertainty increments distributed equally over all *S*s? Changes in the direction of the induction were computed for each *S* and *S*s in each condition were divided at the median into high and low changers. Since the most pronounced individual differences were revealed by the SU measure, analysis was performed using the SU scores. Figure 5 shows mean SU scores for the six trial blocks for the high and low changers.

It is clear from the curves that *S*s who changed their estimates of bomb craters showed considerably more pronounced increments in uncertainty. As a matte of fact, increments in uncertainty obtain only for *S*s who changed their estimates. These results were analyzed by comparing the trial blocks immediately preceding and immediately succeeding stimulus and feedback change. For the purposes of this analysis the results of all change conditions were combined, and the mean SU scores for the two critical trial blocks are shown in Table 6. The analysis of variance for these results is shown in Table 7. It is clear that while the high changers increased their uncertainty, low changers continued to decrease it. It is also clear that the likelihood of change is higher for *S*s with high original uncertainty, although the analysis failed to generate a significant interaction.

This result supports in part the implication of the unfreezing hypothesis, which holds that the unfreezing process occurs only when necessary, that is, when the stability prior to change is relatively high. The differences in uncertainty increments between the b group and the c group also give some support to this implication, although neither of these findings is very convincing. A more direct test is obtained by comparing *S*s whose uncertainty level before change induction was relatively high with those whose uncertainty level was low. Such a comparison was made and the results are shown in Figure 6. By means of the SU scores on the trial block immediately preceding change induction, *S*s in all change con-

TABLE 6

MEAN SU SCORES BEFORE AND AFTER STIMULUS AND FEEDBACK CHANGE FOR HIGH AND LOW CHANGERS

	Mean SU Scores	
	Before Change	After Change
High Changers	101.9	108.5
Low Changers	81.3	71.5

TABLE 7

ANALYSIS OF VARIANCE FOR DATA IN TABLE 6

Source	*df*	*MS*	*F*
High vs. Low Changers (C)	1	29,756	2.16
*S*s in same group	70	13,798	
Trial blocks (B)	1	97	<1.00
B × C	1	2,418	8.17**
Pooled *S*s by B	70	296	

** $p < .01$.

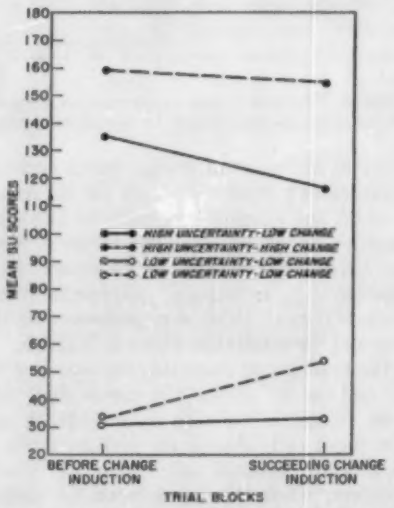

FIG. 6. Effects of cognitive change on uncertainty (SU scores) for subjects with initially high and low uncertainty.

TABLE 8
ANALYSIS OF VARIANCE FOR DATA IN FIGURE 6

Source	*df*	*MS*	*F*
Uncertainty (U)	1	387,195	45.14***
High vs. Low Changers (C)	1	17,009	1.98
U × C	1	3,239	<1.00
Ss in same group	68	8,577	
Trial blocks (T)	1	13	<1.00
T × U	1	5,053	12.86***
T × C	1	2,154	5.48*
T × U × C	1	142	<1.00
Pooled *Ss* × T	68	393	

* $p < .05$.
*** $p < .001$.

ditions were split at the median, and each half was further divided at the median in terms of change in the direction of slide change or feedback change. Thus four groups of *Ss* were obtained; High Uncertainty–High Change, High Uncertainty–Low Change; Low Uncertainty–High Change, and Low Uncertainty–Low Change. The analysis was performed on the uncertainty changes from the trial block immediately preceding to the trial block immediately succeeding change induction, and the analysis of variance is shown in Table 8. It is evident from Figure 6 that increment in uncertainty occurs only for *Ss* whose initial uncertainty was low. *Ss* with high initial uncertainty show decrements whether they changed their estimates or not. The interaction between Trial Blocks and Uncertainty is highly significant. It is also clear that in the Low Uncertainty group there is some increment in uncertainty for *Ss* who changed their estimates relatively less. The previous result relating increments in uncertainty to changes in estimates is again confirmed. In the Low Uncertainty group, those *Ss* who changed their estimates show a considerably more pronounced uncertainty increment than those who did not. As predicted, in the High Uncertainty group there is no increment in uncertainty for high changers, but *Ss* above the median amount of change reduced their uncertainty to a much lesser degree than *Ss* below median amount of change. The interaction between Trial Blocks and Change is significant at the .05 level. The strong main effect of uncertainty is of course due to the arbitrary division of the population at the median.

SUMMARY

The purpose of this study was to examine the role of uncertainty in cognitive change. It was hypothesized that when a person must for some reason change his judgments or opinions his uncertainty should increase. Moreover, it was predicted that uncertainty increments are necessary if cognitive change is to occur, but only when the initial uncertainty is relatively low. Ninety-six subjects were given the task of estimating the number of bomb craters in an aerial photograph in 18 trials involving brief exposures of the stimulus. Judgmental change was induced by either changing the stimulus or changing feedback. Objective and subjective measures of uncertainty were obtained. The results lead to the following conclusions:

1. Repeated exposures of the stimulus result in progressive uncertainty decrements.
2. When requirement for change is introduced the level of uncertainty rises.
3. The extent of the uncertainty increments depends on the uncertainty level before change. With high uncertainty no increments in uncertainty occur; with low uncertainty there are pronounced increments.
4. Uncertainty increments occur only for persons who change their judgments.

REFERENCES

GRANT, D. A. Analysis-of-variance tests in the analysis and comparison of curves. *Psychol. Bull.*, 1956, **53**, 141–154.

LEVITT, E. E. The Water-Jar Einstellung Test as a measure of rigidity. *Psychol. Bull.*, 1956, **53**, 347–370.

LEWIN, K. Group decision and social change. In G. E. Swanson, T. M. Newcomb, & E. L. Hartley, (Eds.), *Readings in Social Psychology*. (Rev. ed.) New York: Holt, 1952. Pp. 459–473.

ZAJONC, R. B., & BURNSTEIN, E. Cognitive behavior under uncertainty. *Amer. Psychologist*, 1957, **12**, 388. (Abstract)

(Received August 17, 1959)

Journal of Abnormal and Social Psychology
1960, Vol. 61, No. 2, 176-180

LEVELING-SHARPENING AND MEMORY ORGANIZATION[1]

PHILIP S. HOLZMAN AND RILEY W. GARDNER
Menninger Foundation, Topeka, Kansas

THE terms, "leveling" and "sharpening," have been used to describe systematic changes in memory organization (Carmichael, Hogan, & Walter, 1932) and the fate of rumors (Allport & Postman, 1947). Omission of inconsistencies, condensation of elements, and general simplification of recalled stories have all been called "leveling." The tendency of a detail of the original structure of a story to dominate its memory organization has been called "sharpening." Wulf (1922) reported analogous processes in the reproduction of geometric figures.

Our use of the terms, "leveling" and "sharpening," shares certain features of these meanings, but implies more. We employ the terms to describe opposite poles of a dimensional principle of cognitive control concerning the degree of assimilation between perceptual processes and memory traces (Köhler, 1923; Lauenstein, 1932). The concept of cognitive control is based on the observation that individual differences in cognitive behaviors often reflect habitual, but not necessarily conscious, strategies of adaptation, and the further observation (see Klein, 1954) that the enduring response dispositions represented by such strategies serve to *mediate* the effects of drives upon cognitive behavior. Cognitive controls such as those we have referred to as the leveling-sharpening dimension are presumed to be structural principles of personality organization that give a particular cast to an individual's cognitive behavior.

Generalized leveling-sharpening tendencies have been explored and defined in several prior studies. The assimilation hypothesis just mentioned led to successful prediction of significant relationships between performance in a laboratory test of assimilation-proneness (the Schematizing Test) and the degree of assimilation shown in visual, auditory, and kinesthetic time error (Holzman, 1954; Holzman & Klein, 1954). Extreme leveling has also been shown to be associated with reliance upon repression as a preferred mode of defense (Gardner, Holzman, Klein, Linton, & Spence, 1959; Holzman & Gardner, 1959) and degree of distortion experienced through aniseikonic lenses (Gardner, et al., 1959).

The most complete study of the leveling-sharpening control principle to date was done by Gardner et al. (1959). In that study, leveling-sharpening emerged as a clear-cut factor for women, and included test scores used as measures of leveling-sharpening in previous studies, among other measures. This factor was described as pertaining "to differentiation in memory organization as a function of the extent to which successive stimuli assimilate to each other." Apparently because of relatively great assimilation among new percepts and relevant memory traces formed earlier, subjects (*S*s) with high factor scores (levelers) formed undifferentiated memory organizations in several tests. *S*s with low factor scores (sharpeners) formed relatively differentiated memory organizations, presumably by virtue of minimal assimilation. A definitive test of the differences in memory schemata in levelers and sharpeners was not, however, undertaken. The present study represents a first such test.

In the earlier studies cited, the criterion of leveling-sharpening has been the Schematizing Test. A complete description of the test may be found in Gardner et al. (1959). Briefly, *S*s are asked to judge in inches the sizes of 150 squares successively projected on a screen. The 14 stimuli range in size from approximately 1 to 14 inches on a side and are projected in a prescribed order that is unknown to the *S*s. The five smallest squares are individually presented (3″ exposure, 8″ interval) in ascending order, then in two random orders. Without *S*'s knowledge, the smallest square is then dropped and one just larger than any seen previously is added. This group of five squares is also presented in ascending order and in two random orders. Subtraction of the

[1] This study was made possible by Research Grant M-1182 from the National Institutes of Health, Public Health Service. We are indebted to our colleague, Richard Siegal, for his help with this study, and to Gardner Murphy, Robert R. Holt, and David Rapaport for their helpful criticisms.

smallest square and addition of a larger square is continued in this fashion until all 14 stimuli have been projected.

Several facets of individual difference are apparent in the Schematizing Test. In the present study, we shall deal with only one of these: the accuracy with which *S*s rank the 150 squares presented (computed in terms of ranking accuracy within each series of 15 judgments, a procedure fully described by Gardner et al.). Accurate intraserial ranking requires continuous comparison of the newly exposed square with those presented earlier. If new percepts and memories of earlier squares undergo relatively great change by virtue of assimilation, intraserial ranking accuracy is notably impaired. Thus, discrimination of the relative sizes of temporarily juxtaposed stimuli may be blurred by assimilation.

If our interpretation of the leveling-sharpening principle is correct, *S*s at different points along this continuum should differ predictably in their recall of much earlier experiences, as well as in their registration of temporal sequences of stimulation in the laboratory. Requiring *S*s to retell a story learned long ago, but not purposefully committed to memory, appears to provide a test of the hypothesis that such memory schemata differ in levelers and sharpeners. We would expect young adults who are levelers to have relatively few correct memories of such a story as "The Pied Piper" and to show evidence of relatively great loss and modification of the over-all structure of the story. This general hypothesis is, of course, difficult to put to adequate test: some *S*s may have heard the story more frequently during childhood than others, some may have reheard the story more recently than others, etc. These features of their prior experience can obviously not be controlled. Control over other extraneous determinants of recall may have been provided, however, by the following facts: (*a*) the age range of our 41 female *S*s was only 18 to 22, (*b*) their socioeconomic backgrounds were remarkably similar, (*c*) they were all sorority members at one small midwestern university. If age, for example, had been allowed to vary more, the relationships reported here might easily have been obscured.

METHOD

The Schematizing Test was administered to 41 *S*s under the standard conditions described by Gardner et al. For purposes of the present study *S*s were divided into groups of 16 levelers and 25 sharpeners on the basis of the mean percentage accuracy of intraserial ranking for five previous samples totaling 175 female *S*s.

In a different testing session, each *S* was asked to tell the story of "The Pied Piper."[2] (Those who reported never having heard it were eliminated.) The stories were recorded.

For purposes of scoring number of elements correct, the story was arbitrarily divided into 11 thematic units:

1. Mention of the town of Hamlin.
2. The town was infested with rats.
3. The townspeople were unable to rid themselves of the rats.
4. The Pied Piper, however, offered to exterminate the rats.
5. If a specified sum of money were paid to him.
6. When he played his pipe, the rats followed him.
7. To a body of water where they drowned.
8. The mayor (or townspeople) refused to pay the Piper.
9. The Piper then played his pipe, which, this time, attracted all the children of the town.
10. The Piper led them out of town into a mountain.
11. Only the lame boy remained behind, because he was unable to keep up with the rest.

Since the story presumably represented "old" memory for most of our *S*s, it seemed advisable to scrutinize the preservation of the main outlines of the story, as well as the correct and incorrect responses at the more microscopic level. The two major sections of the story are linked by a key theme: the Pied Piper's leading the children away as *retribution* for the town fathers' refusal to pay him. Each protocol was thus scored for presence or absence of this key theme, without which ellipsis, condensation, or omission of parts of the over-all organization must occur.

In addition, the following error scores were obtained by the two experimenters (*E*s): number of new elements; number of recombinations; number of fragmentations; number of confabulations, i.e., the number of thematic units associatively elaborated. Because of the simplicity of the scoring scheme, there were only minor differences in scoring. These were easily reconciled to provide a single set of scores.

RESULTS

Quantitative

As anticipated, sharpeners were superior to levelers in their recall of "The Pied Piper." The mean numbers of elements correct were

[2] The use of this story was suggested by David Rapaport.

5.6 for sharpeners, 3.3 for levelers, $t = 2.24$, $p < .02$, one-tailed test.[3]

The two groups also differed in retention of the key "retribution" theme. Only 26% of the levelers retained this theme, whereas 58% of the sharpeners' stories were anchored in this central organizational feature of the story.

The stories told by sharpeners were more than half again as long as those told by levelers. Mean number of typed lines in the recall protocols were: sharpeners, 12.9, levelers, 8.3; $t = 1.86$, $p < .05$, one-tailed test. The two groups were not significantly different in number of new elements, number of recombinations of elements, number of fragmentations, or number of confabulations. The fact that sharpeners produced much longer stories without more errors means, however, that many of their additional elements—those not included in our scoring scheme, which dealt only with correct elements and gross errors—were "sensible" additions that bore a reasonable relationship to the original organization of the story. Since the number-correct score was based only on elements common to various versions of "The Pied Piper," many of these additional elements may represent the particular versions originally heard or read by individual sharpeners.

Qualitative

Qualitative differences in the performances of levelers and sharpeners were perhaps even more impressive than the quantifiable differences just described.

Consider first a performance by one of the levelers that is unique in respect to the core "retribution" theme. This *S* demonstrated her loss of the over-all organization of the story by presenting *two* stories—representing the last and then the first parts of the original story. The loss of the key retribution theme showed itself in a fragmentation of the over-all structure of the story such that *S* apparently remembered it as two separate and unreconcilable entities.

[3] In a large group of which the present *S*s were a part, the Pearson correlations between ACE scores and leveling-sharpening were not significantly different from zero. Intelligence as traditionally measured, then, is not a significant factor in the differences between levelers and sharpeners on recall of "The Pied Piper."

Well there was once a man in this village who played a very wonderful, I guess it was a magic flute, and he could walk through the town and all the children would hear him and be so entranced that they would follow him along to listen to him play. And so he, I guess, went through the villages collecting all the children. I [pause], let's see. . . . Oh . . . I think *this* is the, I think *this* is the Pied Piper. This village was full of mice and rats and so in order to get rid of them they loved music and so there was this man that could play a flute and so he went through the village and played the flute and out of the houses came the mice and rats and they followed him and he led them into the ocean and in that way the town got rid of all the rodents [laughs]. [*E* asks *S* about first story.] Oh, I can't answer the first one because I can't remember what he did with the children, the first story, the one I started and didn't finish.

In the following story, told by a sharpener, the sequence and transition are much more clearly preserved.

The people of Hamlin were over-run by rats and they had tried all sorts of remedies to get rid of the rats and nothing worked. So, one day, they were discussing the rat problem as usual, and this man came to town and the mayor had just previously offered a bag of gold to anyone who could rid the town of rats, so this man came to town. His name was Pied Piper and he told the mayor that he would rid the town of rats for the bag of gold. So, he played on his little pipe and all the rats followed him out—came down from the houses and followed him down the street and took them out of town and they disappeared. He came back for his gold, and the mayor refused to give it to him on the grounds that he never promised it. So Pied Piper told him he would be sorry if he didn't give him the gold, that he would take away all the children. Well, the mayor just laughed at him because he didn't think he could do it. So, he still refused to give him the gold, so the Pied Piper played the tune again and all the children followed him out of the town and he took them out into the country where some mountains—supposedly, the mountains opened up and all the children all went inside to this paradise, and the children—the mountains closed and the children were never seen again.

Vagueness was also notable in levelers' recall. Often only an indistinct general impression of the story could be recovered, with important details missing. The following story by a leveler seems to highlight these qualities.

Well, I probably didn't know any facts about it but, well, let's see, didn't he get rid of the mice or rats or something from the town? I don't remember. But as I remember, they, they were, they were trying to get rid of them or something, and he by playing his pipe, or whatever it was, musical instrument, anyway, by doing that, he drove them out of town, but that's all I've ever known. [*E*: Do you remember anything else

at all?] *S*: Unh uh. [*E*: Do you have the feeling that there's more to the story?] *S*: Oh, yes, but I probably just don't know.

DISCUSSION

Theories of learning and remembering may be arranged along a continuum of greater to lesser influence of "personality" factors on these cognitive functions. In the special theories of Ebbinghaus, for example, the role of general personality factors enters not at all. Associationistic theories of learning in which the law of effect plays a crucial role emphasize the part played by affective factors in learning and retention. Gestalt theories of learning and remembering elucidate chiefly the ways in which structural properties of the material to be learned determine registration and recall. Others, principally Bartlett (1932), have refined our awareness that memories themselves have organizational properties shaped in part by affects. Psychoanalysis has attempted to take into account the central role of affects in memory in the theory of repression. A complete review and thoughtful critique of these issues has been provided by Rapaport (1950). It may fairly be said that, in general, conceptions of the role of personality in memory have been limited to special effects, such as needs, affects, and specific sets. The concentration on such influences upon memory seems to emerge from an implicit assumption that personality factors are external influences on the learning or recall processes, factors that act either to enhance or to interfere with memory.

Theories of learning and remembering have usually been constructed without reference to individual differences in personality *organization* that shape both registration and recall. The general laws propounded thus far bypass the possibility that Person A and Person B may learn and remember in strikingly different ways, although when their performances are taken together with those of others, the group averages may show conformity to such laws.[4] Recall is a complex process in which details lose or maintain their place within the context of the structural organization of the story as this structure was shaped at the time(s) of the person's exposure to the story.

The cognitive control principles referred to earlier are concepts employed to describe and account for some of the major dimensions along which *personal constants* in thinking, perceiving, remembering, etc., fall. These controls are assumed to represent aspects of the individual's personality organization that determine the nature and quality of such cognitive functions. Leveling-sharpening—a dimension of consistent individual differences in susceptibility to assimilation (process-trace interaction)—is a principle of cognitive organization that affects the registration, and hence recall, of stimuli in a wide variety of situations, here shown to include a test of recall of "The Pied Piper" story. Leveling-sharpening thus seems to be apparent in the organization of old memories, as well as in the organization of memories registered in a current laboratory experiment. This finding implies that leveling-sharpening is a generalized and enduring aspect of cognitive organization. *S*s at the leveling end of this dimension of cognitive organization formed memory schemata for a story heard much earlier that have the following major characteristics: loss of the over-all structure of the story, loss of elements as a product of "fusion" in the process of registration, relatively large numbers of elements (considering the number of recallable elements available) contaminated or changed by interaction with other memory-elements-to-be during the registration process.

Repression, one of the special "personality" factors affecting memory, has received considerable clinical attention from psychoanalysts. In previous studies (Gardner, et al., 1959; Holzman & Gardner, 1959), we showed that persons who employ repression as a principal defense mechanism tend to be levelers, i.e., that a major characteristic of the memory apparatus is discernible in both defensive and adaptive uses of memory. In the light of the earlier findings, results of the present study also suggest that a generalized tendency to form undifferentiated memory schemata provides a necessary but not a sufficient condition for the unavailability to intentional verbal expression of ideas that, in psychoanalytic

[4] I. H. Paul's (1959) study of individual differences in remembering, however, represents a novel and fruitful approach to understanding individual styles of learning and recall, particularly as these styles affect the "importation" of various kinds of extraneous material in recall.

terms, have become drive-cathected. Such generalized clogging in the recall of ideas by levelers is consistent with McGeoch's (1929) demonstration that the higher the degree of the original learning, the less the interference in the reproduction of the material. A further investigation is indicated to determine whether, among levelers, repression occurs in a diffuse fashion, while instances of repression among sharpeners are more discrete and confined to special circumstances.

Experiments now in progress are designed to determine whether or not levelers' undifferentiated initial memory organizations are subject to relatively great "spontaneous" changes after original registration is completed. Although the present study points to a link between registration in the Schematizing Test, earlier registration of "The Pied Piper" story, and subsequent registration of events *S* experienced as related to the story (at some level of conceptual organization), our results may also be due to post-registration changes that result primarily from lack of differentiation in levelers' memories. A further possibility, pointed to by Gardner et al. (1959), must also be put to the test: that proneness to assimilative phenomena is a function of the intensity with which persons attend to new experiences. Persons who characteristically "cathect" new experiences weakly may be susceptible to assimilation (i.e., may be levelers) because of vagueness and instability in their original percepts.

REFERENCES

ALLPORT, G. W., & POSTMAN, L. *The psychology of rumor*. New York: Holt, 1947.

BARTLETT, F. C. *Remembering: A study in experimental and social psychology*. Cambridge: Cambridge Univer. Press, 1932.

CARMICHAEL, L., HOGAN, H. P., & WALTER, A. A. An experimental study of the effect of language on the reproduction of visually perceived form. *J. exp. Psychol.*, 1932, **15**, 73–86.

GARDNER, R. W., HOLZMAN, P. S., KLEIN, G. S., LINTON, HARRIET B., & SPENCE, D. P. Cognitive control: A study of individual consistencies in cognitive behavior. *Psychol. Issues*, 1959, **1**, No. 4.

HOLZMAN, P. S. The relation of assimilation tendencies in visual, auditory, and kinesthetic time-error to cognitive attitudes of leveling and sharpening. *J. Pers.*, 1954, **22**, 375–394.

HOLZMAN, P. S., & GARDNER, R. W. Leveling and repression. *J. abnorm. soc. Psychol.*, 1959, **59**, 151–155.

HOLZMAN, P. S., & KLEIN, G. S. Cognitive system-principles of leveling and sharpening: Individual differences in assimilation effects in visual time-error. *J. Psychol.*, 1954, **37**, 105–122.

KLEIN, G. S. Need and regulation. In M. R. Jones (Ed.), *Nebraska symposium on motivation*. Lincoln: Univer. Nebraska Press, 1954. Pp. 224–274.

KÖHLER, W. Zur Theorie des Sukzessivvergleichs und der Zeitfehler. *Psychol. Forsch.*, 1923, **4**, 115–175.

LAUENSTEIN, O. Ansatz zu einer physiologischen Theorie des Vergleichs und der Zeitfehler. *Psychol. Forsch.*, 1932, **17**, 130–177.

MCGEOCH, J. A. The influence of degree of learning upon retroactive inhibition. *Amer. J. Psychol.*, 1929, **41**, 252–262.

PAUL, I. H. Studies in remembering: The reproduction of connected and extended verbal material. *Psychol. Issues*, 1959, **1**, No. 2.

RAPAPORT, D. *Emotions and memory*. New York: International Univer. Press, 1950.

WULF, F. Über die Veränderung von Vorstellungen (Gedächtnis und Gestalt). *Psychol. Forsch.*, 1922, **1**, 333–373.

(Early publication received April 4, 1960)

Journal of Abnormal and Social Psychology
1960, Vol. 61, No. 2, 181–188

THE EFFECT OF THREAT UPON INTERPERSONAL BARGAINING

MORTON DEUTSCH AND ROBERT M. KRAUSS
Bell Telephone Laboratories, Incorporated, Murray Hill, New Jersey

A BARGAIN is defined in *Webster's Unabridged Dictionary* as "an agreement between parties settling what each shall give and receive in a transaction between them"; it is further specified that a bargain is "an agreement or compact viewed as advantageous or the reverse." When the term "agreement" is broadened to include tacit, informal agreements as well as explicit agreements, it is evident that bargains and the processes involved in arriving at bargains ("bargaining") are pervasive characteristics of social life.

The definition of bargain fits under sociological definitions of the term "social norm." In this light, the experimental study of the bargaining process and of bargaining outcomes provides a means for the laboratory study of the development of certain types of social norms. But unlike many other types of social situations, bargaining situations have certain distinctive features that make it relevant to consider the conditions that determine whether or not a social norm will develop as well as those that determine the nature of the social norm if it develops. Bargaining situations highlight the possibility that, even where cooperation would be mutally advantageous, shared purposes may not develop, agreement may not be reached, and interaction may be regulated antagonistically rather than normatively.

The essential features of a bargaining situation exist when:

1. Both parties perceive that there is the possibility of reaching an agreement in which each party would be better off, or no worse off, than if no agreement were reached.

2. Both parties perceive that there is more than one such agreement that could be reached.

3. Both parties perceive each other to have conflicting preferences or opposed interests with regard to the different agreements that might be reached.

Everyday examples of bargaining include such situations as: the buyer-seller relationship when the price is not fixed, the husband and wife who want to spend an evening out together but have conflicting preferences about where to go, union-management negotiations, drivers who meet at an intersection when there is no clear right of way, disarmament negotiations.

In terms of our prior conceptualization of cooperation and competition (Deutsch, 1949) bargaining is thus a situation in which the participants have mixed motives toward one another: on the one hand, each has interest in cooperating so that they reach an agreement; on the other hand, they have competitive interests concerning the nature of the agreement they reach. In effect, to reach agreement the cooperative interest of the bargainers must be strong enough to overcome their competitive interests. However, agreement is not only contingent upon the *motivational* balances of cooperative to competitive interests but also upon the situational and *cognitive* factors which facilitate or hinder the recognition or invention of a bargaining agreement that reduces the opposition of interest and enhances the mutuality of interest.[1]

These considerations lead to the formulation of two general, closely related propositions about the likelihood that a bargaining agreement will be reached.

1. Bargainers are more likely to reach an agreement, the stronger are their cooperative interests in comparison with their competitive interests.

2. Bargainers are more likely to reach an agreement, the more resources they have available for recognizing or inventing potential bargaining agreements and for communicating to one another once a potential agreement has been recognized or invented.

From these two basic propositions and additional hypotheses concerning conditions that determine the strengths of the cooperative and competitive interests and the amount of available resources, we believe it is possible to explain the ease or difficulty of arriving at a bar-

[1] Schelling in a series of stimulating papers on bargaining (1957, 1958) has also stressed the "mixed motive" character of bargaining situations and has analyzed some of the cognitive factors which determine agreements.

gaining agreement. We shall not present a full statement of these hypotheses here but turn instead to a description of an experiment that relates to Proposition 1.

The experiment was concerned with the effect of the availability of threat upon bargaining in a two-person experimental bargaining game.[2] Threat is defined as the expression of an intention to do something detrimental to the interests of another. Our experiment was guided by two assumptions about threat:

1. If there is a conflict of interest and one person is able to threaten the other, he will tend to use the threat in an attempt to force the other person to yield. This tendency should be stronger, the more irreconcilable the conflict is perceived to be.

2. If a person uses threat in an attempt to intimidate another, the threatened person (if he considers himself to be of equal or superior status) would feel hostility toward the threatener and tend to respond with counterthreat and/or increased resistance to yielding. We qualify this assumption by stating that the tendency to resist should be greater, the greater the perceived probability and magnitude of detriment to the other and the less the perceived probability and magnitude of detriment to the potential resister from the anticipated resistance to yielding.

The second assumption is based upon the view that when resistance is not seen to be suicidal or useless, to allow oneself to be intimidated, particularly by someone who does not have the right to expect deferential behavior, is to suffer a loss of social face and, hence, of self-esteem; and that the culturally defined way of maintaining self-esteem in the face of attempted intimidation is to engage in a contest for supremacy vis-à-vis the power to intimidate or, minimally, to resist intimidation. Thus, in effect, the use of threat (and if it is available to be used, there will be a tendency to use it) should strengthen the competitive interests of the bargainers in relationship to one another by introducing or enhancing the competitive struggle for self-esteem. Hence, from Proposition 1, it follows that the availability of a means of threat should make it more difficult for the bargainers to reach agreement (providing that the threatened person has some means of resisting the threat). The preceding statement is relevant to the comparison of both of our experimental conditions of threat, bilateral and unilateral (described below), with our experimental condition of nonthreat. We hypothesize that a bargaining agreement is more likely to be achieved when neither party can threaten the other, than when one or both parties can threaten the other.

Consider now the situations of bilateral threat and unilateral threat. For several reasons, a situation of bilateral threat is probably less conducive to agreement than is a condition of unilateral threat. First, the sheer likelihood that a threat will be made is greater when two people rather than one have the means of making the threat. Secondly, once a threat is made in the bilateral case it is likely to evoke counterthreat. Withdrawal of threat in the face of counterthreat probably involves more loss of face (for reasons analogous to those discussed in relation to yielding to intimidation) than does withdrawal of threat in the face of resistance to threat. Finally, in the unilateral case, although the person without the threat potential can resist and not yield to the threat, his position vis-à-vis the other is not so strong as the position of the threatened person in the bilateral case. In the unilateral case, the threatened person may have a worse outcome than the other whether he resists or yields; while in the bilateral case, the threatened person is sure to have a worse outcome if he yields but he may insure that he does not have a worse outcome if he does not yield.

METHOD

Procedure

Subjects (*Ss*) were asked to imagine that they were in charge of a trucking company, carrying merchandise over a road to a destination. For each trip completed they made $.60, minus their operating expenses. Operating expenses were calculated at the rate of one cent per second. So, for example, if it took 37 seconds to complete a particular trip, the player's profit would be $.60 − $.37 or a net profit of $.23 for that particular trip.

Each *S* was assigned a name, Acme or Bolt. As the "road map" (see Figure 1) indicates, both players start from separate points and go to separate destinations. At one point their paths cross. This is the section of road labeled "one lane road," which is only one lane wide, so that two trucks, heading in opposite directions, could not pass each other. If one backs up

[2] The game was conceived and originated by M. Deutsch; R. M. Krauss designed and constructed the apparatus employed in the experiment.

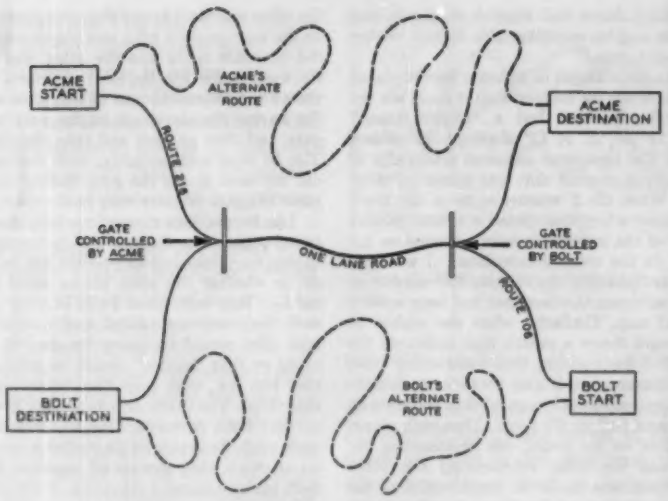

FIG. 1. Subject's road map.

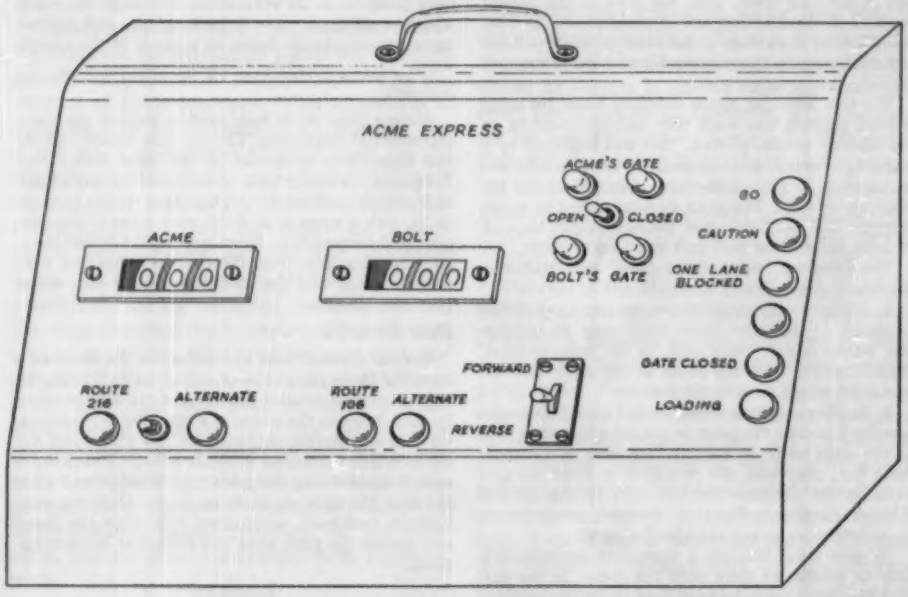

FIG. 2. Subject's control panel.

the other can go forward, or both can back up, or both can sit there head-on without moving.

There is another way for each *S* to reach the destination on the map, labeled the "alternate route." The two players' paths do not cross on this route, but the alternate is 56% longer than the main route. *S*s were told that they could expect to lose at least $.10 each time they used the alternate route.

At either end of the one-lane section there is a gate that is under the control of the player to whose starting point it is closest. By closing the gate, one player can prevent the other from traveling over that section of the main route. The use of the gate provides the threat potential in this game. In the bilateral threat potential condition (Two Gates) both players had gates under their control. In a second condition of unilateral

threat (One Gate) Acme had control of a gate but Bolt did not. In a third condition (No Gates) neither player controlled a gate.

Ss played the game seated in separate booths placed so that they could not see each other but could see the experimenter (*E*). Each *S* had a "control panel" mounted on a 12″ × 18″ × 12″ sloping-front cabinet (see Figure 2). The apparatus consisted essentially of a reversible impulse counter that was pulsed by a recycling timer. When the *S* wanted to move her truck forward she threw a key that closed a circuit pulsing the "add" coil of the impulse counter mounted on her control panel. As the counter cumulated, *S* was able to determine her "position" by relating the number on her counter to reference numbers that had been written in on her road map. Similarly, when she wished to reverse, she would throw a switch that activated the "subtract" coil of her counter, thus subtracting from the total on the counter each time the timer cycled.

S's counter was connected in parallel to counters on the other *S*'s panel and on *E*'s panel. Thus each player had two counters on her panel, one representing her own position and the other representing the other player's. Provision was made in construction of the apparatus to permit cutting the other player's counter out of the circuit, so that each *S* knew only the position of her own truck. This was done in the present experiment. Experiments now in progress are studying the effects of knowledge of the other person's position and other aspects of interpersonal communication upon the bargaining process.

The only time one player definitely knew the other player's position was when they had met head-on on the one-way section of road. This was indicated by a traffic light mounted on the panel. When this light was on, neither player could move forward unless the other moved back. The gates were controlled by toggle switches and panel-mounted indicator lights showed, for both *Ss*, whether each gate was open or closed.

The following "rules of the game" were stated to the *Ss*:

1. A player who started out on one route and wished to switch to the other route could only do so after first reversing and going back to the start position. Direct transfer from one route to the other was not permitted except at the start position.

2. In the conditions where *Ss* had gates, they were permitted to close the gates no matter where they were on the main route, so long as they were on the main route (i.e., they were not permitted to close the gate while on the alternate route or after having reached their destinations). However, *Ss* were permitted to open their gates at any point in the game.

Ss were taken through a number of practice exercises to familiarize them with the game. In the first trial they were made to meet head-on on the one-lane path; Acme was then told to back up until she was just off the one-lane path and Bolt was told to go forward. After Bolt had gone through the one-lane path, Acme was told to go forward. Each continued going forward until each arrived at her destination. The second practice trial was the same as the first except that Bolt rather than Acme backed up after meeting head-on. In the next practice trial, one of the players was made to wait just before the one-way path while the other traversed it and then was allowed to continue. In the next practice trial, one player was made to take the alternate route and the other was made to take the main route. Finally, in the bilateral and unilateral threat conditions the use of the gate was illustrated (by having the player get on the main route, close the gate, and then go back and take the alternate route). The *Ss* were told explicitly, with emphasis, that they did *not* have to use the gate. Before each trial in the game the gate or gates were in the open position.

The instructions stressed an individualistic motivational orientation. *Ss* were told to try to earn as much money for themselves as possible and to have no interest in whether the other player made money or lost money. They were given $4.00 in poker chips to represent their working capital and told that after each trial they would be given "money" if they made a profit or that "money" would be taken from them if they lost (i.e., took more than 60 seconds to complete their trip). The profit or loss of each *S* was announced so that both *Ss* could hear the announcement after each trial. Each pair of *Ss* played a total of 20 trials; on all trials, they started off together. In other words each trial presented a repetition of the same bargaining problem. In cases where *Ss* lost their working capital before the 20 trials were completed, additional chips were given them. *Ss* were aware that their monetary winnings and losses were to be imaginary and that no money would change hands as a result of the experiment.

Subjects

Sixteen pairs of *Ss* were used in each of the three experimental conditions. The *Ss* were female clerical and supervisory personnel of the New Jersey Bell Telephone Company who volunteered to participate during their working day.[3] Their ages ranged from 20 to 39, with a mean of 26.2. All were naive to the purpose of the experiment. By staggering the arrival times and choosing girls from different locations, we were able to insure that the *Ss* did not know with whom they were playing.

Data Recorded

Several types of data were collected. We obtained a record of the profit or loss of each *S* on each trial. We also obtained a detailed recording of the actions taken by each *S* during the course of a trial. For this purpose, we used an Esterline-Angus model AW Operations Recorder which enabled us to obtain a "log" of each move each *S* made during the game (e.g., whether and when she took the main or alternate route; when she went forward, backward, or remained still; when she closed and opened the gate; when she arrived at her destination).

RESULTS[4]

The best single measure of the difficulty experienced by the bargainers in reaching an agreement is the sum of each pair's profits (or

[3] We are indebted to the New Jersey Bell Telephone Company for their cooperation in providing *Ss* and facilities for the experiment.

[4] We are indebted to M. J. R. Healy for suggestions concerning the statistical analysis of our data.

losses) on a given trial. The higher the sum of the payoffs to the two players on a given trial, the less time it took them to arrive at a procedure for sharing the one-lane path of the main route. (It was, of course, possible for one or both of the players to decide to take the alternate route so as to avoid a protracted stalemate during the process of bargaining. This, however, always resulted in at least a $.20 smaller joint payoff if only one player took the alternate route, than an optimally arrived at agreement concerning the use of the one-way path.) Figure 3 presents the medians of the summed payoffs (i.e., Acme's plus Bolt's) for all pairs in each of the three experimental conditions over the 20 trials.[5] These striking results indicate that agreement was least difficult to arrive at in the no threat condition, was more difficult to arrive at in the unilateral threat condition, and exceedingly difficult or impossible to arrive at in the bilateral threat condition (see also Table 1).

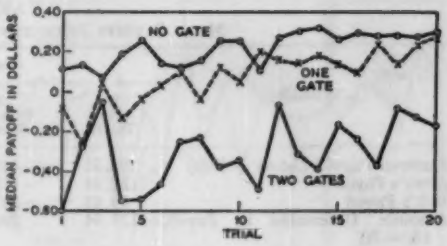

FIG. 3. Median joint payoff (Acme + Bolt) over trials.

Examination of Figure 3 suggests that learning occurred during the 20 trials: the summed payoffs for pairs of *S*s tend to improve as the number of trials increases. This suggestion is confirmed by an analysis of variance of the slopes for the summed payoffs[6] over the 20 trials for each of the 16 pairs in each of the 3 experimental treatments. The results of this analysis indicate that the slopes are significantly greater than zero for the unilateral threat ($p < .01$) and the no threat ($p < .02$) conditions; for the bilateral threat condition, the slope does not reach statistical significance ($.10 < p < .20$). The data indicate that the pairs in the no threat condition started off at a fairly high level but, even so, showed some improvement over the 20 trials; the pairs in the unilateral threat condition started off low and, having considerable opportunity for improvement, used their opportunity; the pairs in the bilateral threat condition, on the other hand, did not benefit markedly from repeated trials.

Figure 4 compares Acme's median profit in the three experimental conditions over the 20 trials; while Figure 5 compares Bolt's profit in the three conditions. (In the unilateral threat condition, it was Acme who controlled a gate and Bolt who did not.) Bolt's as well as Acme's outcome is somewhat better in the no threat condition than in the unilateral threat condition; Acme's, as well as Bolt's, outcome is clearly worst in the bilateral threat condition (see Table 1 also). However, Figure 6 reveals that Acme does somewhat better than Bolt in the unilateral condition. Thus, if threat-potential exists within a bargaining relationship it is better to possess it oneself than to have the other party possess it. However, it is even better for neither party to possess it. Moreover, Figure 5 shows that Bolt is better off not having than having a gate even when Acme has a gate: Bolt tends to do better in the unilateral threat condition than in the bilateral threat condition.

The size of the absolute discrepancy between the payoffs of the two players in each pair provides a measure of the confusion or difficulty in predicting what the other player was going to do. Thus, a large absolute discrepancy might indicate that after one player had gone through the one-way path and left it open, the other player continued to wait; or it might indicate that one player continued to wait at a closed gate hoping the other player would open it quickly but the other player did not; etc. Figure 7 indicates that the discrepancy between players in the no threat condition is initially small and remains small for the 20 trials. For the players in both the bilateral and unilateral threat conditions, the discrepancy is initially relatively larger; but it decreases more noticeably in the unilateral threat condition by the tenth trial and, therefore, is consistently smaller than in the bilateral condition.

By way of concrete illustration, we present a

[5] Medians are used in graphic presentation of our results because the wide variability of means makes inspection cumbersome.

[6] A logarithmic transformation of the summed payoffs on each trial for each pair was made before computing the slopes for a given pair.

TABLE 1
MEAN PAYOFFS SUMMATED OVER THE TWENTY TRIALS

Variable	Means			Statistical Comparisons: p values[a]			
	(1) No Threat	(2) Unilateral Threat	(3) Bilateral Threat	Overall	(1)vs. (2)	(1)vs.(3)	(2)vs.(3)
Summed Payoffs (Acme + Bolt)	203.31	−405.88	−875.12	.01	.01	.01	.05
Acme's Payoff	122.44	−118.56	−406.56	.01	.10	.01	.05
Bolt's Payoff	80.88	−287.31	−468.56	.01	.01	.01	.20
Absolute Differences in Payoff (A − B)	125.94	294.75	315.25	.05	.05	.01	ns

[a] Evaluation of the significance of overall variation between conditions is based on an F test with 2 and 45 df. Comparisons between treatments are based on a two-tailed t test.

synopsis of the game for one pair in each of three experimental treatments.

No Threat Condition

Trial 1. The players met in the center of the one-way section. After some back-and-forth movement Bolt reversed to the end of the one-way section, allowing Acme to pass through, and then proceeded forward herself.

Trial 2. They again met at the center of the one-way path. This time, after moving back and forth deadlocked for some time, Bolt reversed to "start" and took the alternate route to her destination, thus leaving Acme free to go through on the main route.

Trial 3. The players again met at the center of the one-way path. This time, however, Acme reversed to the beginning of the path, allowing Bolt to go through to her destination. Then Acme was able to proceed forward on the main route.

Trial 5. Both players elected to take the alternate route to their destinations.

Trial 7. Both players took the main route and met in the center. They waited, deadlocked, for a considerable time. Then Acme reversed to the end of the one-way path allowing Bolt to go through, then proceeded through to her destination.

Trials 10–20. Acme and Bolt fall into a pattern of alternating who is to go first on the one-way section. There is no deviation from this pattern.

The only other pattern that emerges in this condition is one in which one player dominates the other. That is, one player consistently goes first on the one-way section and the other player consistently yields.

Unilateral Threat Condition

Trial 1. Both players took the main route and met in the center of it. Acme immediately closed the gate, reversed to "start," and took the alternate route to her destination. Bolt waited for a few seconds, at the closed gate, then reversed and took the alternate route.

Trial 2. Both players took the main route and met in the center. After moving back and forth deadlocked for about 15 seconds, Bolt reversed to the beginning of the one-way path, allowed Acme to pass, and then proceeded forward to her destination.

Trial 3. Both players started out on the main route, meeting in the center. After moving back and forth deadlocked for a while, Acme closed her gate, reversed to "start," and took the alternate route. Bolt, meanwhile, waited at the closed gate. When Acme arrived at her destination she opened the gate, and Bolt went through to complete her trip.

Trial 5. Both players took the main route, meeting at the center of the one-way section. Acme immediately closed her gate, reversed, and took the alternate route. Bolt waited at the gate for about 10 seconds, then reversed and took the alternate route to her destination.

Trial 10. Both players took the main route and met in the center. Acme closed her gate, reversed, and took the alternate route. Bolt remained waiting at the closed gate. After Acme arrived at her destination, she opened the gate and Bolt completed her trip.

Trial 15. Acme took the main route to her destination and Bolt took the alternate route.

Trials 17–20. Both players took the main route and met in the center. Bolt waited a few seconds, then reversed to the end of the one-way section allowing Acme to go through. Then Bolt proceeded forward to her destination.

Other typical patterns that developed in this experimental condition included an alternating pattern similar to that described in the no threat condition, a dominating pattern in which Bolt would select the alternate route leaving Acme free to use the main route unobstructed, and a pattern in which Acme would close her gate and then take the alternate route, also forcing Bolt to take the alternate route.

Bilateral Threat Condition

Trial 1. Acme took the main route and Bolt took the alternate route.

Trial 2. Both players took the main route and met head-on. Bolt closed her gate. Acme waited a few seconds, then closed her gate, reversed to "start," then went forward again to the closed gate. Acme reversed and took the alternate route. Bolt again reversed, then started on the alternate route. Acme opened her gate and Bolt reversed to "start" and went to her destination on the main route.

Trial 3. Acme took the alternate route to her destination. Bolt took the main route and closed her gate before entering the one-way section.

Trial 5. Both players took the main route and met head-on. After about 10 seconds spent backing up and going forward, Acme closed her gate, reversed, and took the alternate route. After waiting a few seconds, Bolt did the same.

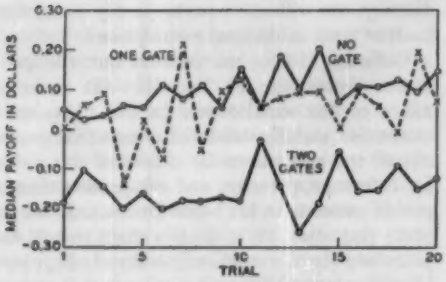

FIG. 4. Acme's median payoff.

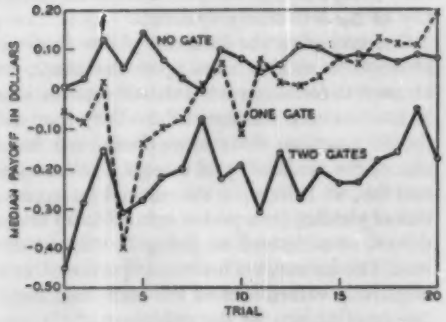

FIG. 5. Bolt's median payoff.

Trials 8–10. Both players started out on the main route, immediately closed their gates, reversed to "start," and took the alternate route to their destinations.

Trial 15. Both players started out on the main route and met head-on. After some jockeying for position, Acme closed her gate, reversed, and took the alternate route to her destination. After waiting at the gate for a few seconds, Bolt reversed to "start" and took the alternate route to her destination.

Trials 19–20. Both players started out on the main route, immediately closed their gates, reversed to "start," and took the alternate routes to their destinations.

Other patterns that emerged in the bilateral threat condition included alternating first use of the one-way section, one player's dominating the other on first use of the one-way section, and another dominating pattern in which one player consistently took the main route while the other consistently took the alternate route.

DISCUSSION

From our view of bargaining as a situation in which both cooperative and competitive tendencies are present and acting upon the individual, it is relevant to inquire as to the conditions under which a stable agreement of any form develops. However, implicit in most economic models of bargaining (e.g., Stone, 1958; Zeuthen, 1930) is the assumption that

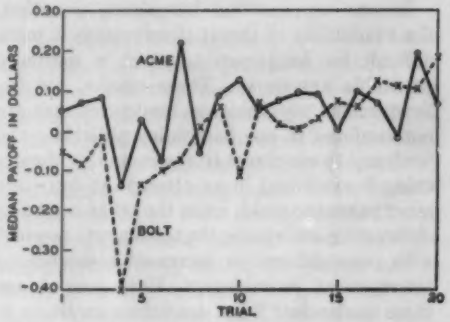

FIG. 6. Acme's and Bolt's median payoffs in unilateral threat condition.

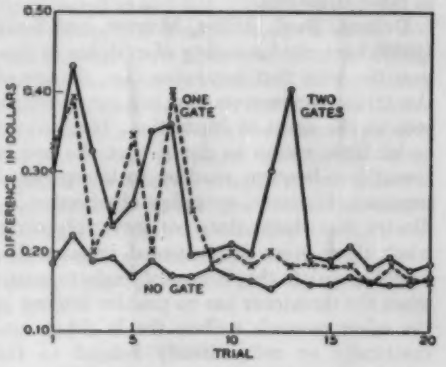

FIG. 7. Median absolute differences in payoff.

the cooperative interests of the bargainers are sufficiently strong to insure that some form of mutually satisfactory agreement will be reached. For this reason, such models have focused upon the form of the agreement reached by the bargainers. Siegel and Fouraker (1960) report a series of bargaining experiments quite different in structure from ours in which only one of many pairs of *S*s were unable to reach agreement. Siegel and Fouraker explain this rather startling result as follows:

> Apparently the disruptive forces which lead to the rupture of some negotiations were at least partially controlled in our sessions . . . .
>
> Some negotiations collapse when one party becomes incensed at the other, and henceforth strives to maximize his opponent's displeasure rather than his own satisfaction . . . . Since it is difficult to transmit insults by means of quantitative bids, such disequilibrating behavior was not induced in the present studies. If subjects were allowed more latitude in their communications and interactions, the possibility of an affront-offense-punitive behavior sequence might be increased (p. 100). (Quoted by permission of McGraw-Hill)

In our experimental bargaining situation, the availability of threat clearly made it more difficult for bargainers to reach a mutually profitable agreement. These results, we believe, reflect psychological tendencies that are not confined to our bargaining situation: the tendency to use threat (if the means for threatening is available) in an attempt to force the other person to yield, when the other is seen as obstructing one's path; the tendency to respond with counterthreat or increased resistance to attempts at intimidation. How general are these tendencies? What conditions are likely to elicit them? Answers to these questions are necessary before our results can be generalized to other situations.

Dollard, Doob, Miller, Mowrer, and Sears (1939) have cited a variety of evidence to support the view that aggression (i.e., the use of threat) is a common reaction to a person who is seen as the agent of frustration. There seems to be little reason to doubt that the use of threat is a frequent reaction to interpersonal impasses. However, everyday observation indicates that threat does not inevitably occur when there is an interpersonal impasse. We would speculate that it is most likely to occur: when the threatener has no positive interest in the other person's welfare (he is either egocentrically or competitively related to the other); when the threatener believes that the other has no positive interest in his welfare; and when the threatener anticipates either that his threat will be effective or, if ineffective, will not worsen his situation because he expects the worst to happen if he does not use his threat. We suggest that these conditions were operative in our experiment; *S*s were either egocentrically or competitively oriented to one another[7] and they felt that they would not be worse off by the use of threat.

Everyday observation suggests that the tendency to respond with counterthreat or increased resistance to attempts at intimidation is also a common occurrence. We believe that introducing threat into a bargaining situation affects the meaning of yielding. Although we have no data to support this interpretation directly, we will attempt to justify it on the basis of some additional assumptions.

Goffman (1955) has pointed out the pervasive significance of "face" in the maintenance of the social order. In this view, self-esteem is a socially validated system that grows out of the acceptance by others of the claim for deference, prestige, and recognition that a person presents in his behavior toward others. Since the rejection of such a claim would be perceived (by the recipient) as directed against his self-esteem, he must react against it rather than accept it in order to maintain the integrity of his self-esteem system.

One may view the behavior of our *S*s as an attempt to make claims upon the other, an attempt to develop a set of shared expectations as to what each was entitled to. Why then did the *S*s' reactions differ so markedly as a function of the availability of threat? The explanation lies, we believe, in the cultural interpretation of yielding (to a peer or subordinate) under duress, as compared to giving in without duress. The former, we believe, is perceived as a negatively valued form of behavior, with negative implications for the self-image of the person who so behaves. At least partly, this is so because the locus of causality is perceived to be outside the person's voluntary control. No such evaluation, however, need be placed on the behavior of one who "gives in" in a situation where no threat or duress is a factor. Rather, we should expect the culturally defined evaluation of such a person's behavior to be one of "reasonableness" or "maturity," because the source of the individual's behavior is perceived to lie within his own control.

Our discussion so far has suggested that the psychological factors which operate in our experimental bargaining situation are to be found in many real-life bargaining situations. However, it is well to recognize some unique features of our experimental game. First, the bargainers had no opportunity to communicate verbally with one another. Prior research on the role of communication in trust (Deutsch 1958, 1960; Loomis, 1959) suggests that the opportunity for communication would have made reaching an agreement easier for individualistically-oriented bargainers. This same research (Deutsch, 1960) indicates, however, that communication may not be effective between competitively oriented bargainers. This

[7] A post-experimental questionnaire indicated that, in all three experimental conditions, the *S*s were most strongly motivated to win money, next most strongly motivated to do better than the other player, next most motivated to "have fun," and were very little or not at all motivated to help the other player.

possibility was expressed spontaneously by a number of our *Ss* in a post-game interview.

Another characteristic of our bargaining game is that the passage of time, without coming to an agreement, is costly to the players. There are, of course, bargaining situations in which lack of agreement may simply preserve the *status quo* without any worsening of the bargainers' respective situations. This is the case in the typical bilateral monopoly case, where the buyer and seller are unable to agree upon a price (e.g., see Siegel & Fouraker, 1960). In other sorts of bargaining situations, however, (e.g., labor-management negotiations during a strike, internation negotiations during an expensive cold war) the passage of time may play an important role. In our experiment, we received the impression that the meaning of time changed as time passed without the bargainers reaching an agreement. Initially, the passage of time seemed to place the players under pressure to come to an agreement before their costs mounted sufficiently to destroy their profit. With the continued passage of time, however, their mounting losses strengthened their resolution not to yield to the other player. They comment: "I've lost so much, I'll be damned if I give in now. At least I'll have the satisfaction of doing better than she does." The mounting losses and continued deadlock seemed to change the game from a mixed motive into a predominantly competitive situation.

It is, of course, hazardous to generalize from a laboratory experiment to the complex problems of the real world. But our experiment and the theoretical ideas underlying it can perhaps serve to emphasize some notions which, otherwise, have an intrinsic plausibility. In brief, these are that there is more safety in cooperative than in competitive coexistence, that it is dangerous for bargainers to have weapons, and that it is possibly even more dangerous for a bargainer to have the capacity to retaliate in kind than not to have this capacity when the other bargainer has a weapon. This last statement assumes that the one who yields has more of his values preserved by accepting the agreement preferred by the other than by extended conflict. Of course, in some bargaining situations in the real world, the loss incurred by yielding may exceed the losses due to extended conflict.

SUMMARY

The nature of bargaining situations was discussed. Two general propositions about the conditions affecting the likelihood of a bargaining agreement were presented. The effects of the availability of threat upon interpersonal bargaining were investigated experimentally in a two-person bargaining game. Three experimental conditions were employed: no threat (neither player could threaten the other), unilateral threat (only one of the players had a means of threat available to her), and bilateral threat (both players could threaten each other). The results indicated that the difficulty in reaching an agreement and the amount of (imaginary) money lost, individually as well as collectively, was greatest in the bilateral and next greatest in the unilateral threat condition. Only in the no threat condition did the players make an overall profit. In the unilateral threat condition, the player with the threat capability did better than the player without the threat capability. However, comparing the bilateral and unilateral threat conditions, the results also indicate that when facing a player who had threat capability one was better off *not* having than having the capacity to retaliate in kind.

REFERENCES

DEUTSCH, M. A theory of cooperation and competition. *Hum. Relat.*, 1949, **2**, 129–152.

DEUTSCH, M. Trust and suspicion. *J. conflict Resolut.*, 1958, **2**, 265–279.

DEUTSCH, M. The effect of motivational orientation upon trust and suspicion. *Hum. Relat.*, 1960, **13**, 123–140.

DOLLARD, J., DOOB, L. W., MILLER, N. E., MOWRER, O. H., & SEARS, R. H. *Frustration and aggression*. New Haven: Yale Univer. Press, 1939.

GOFFMAN, E. On face-work. *Psychiatry*, 1955, **18**, 213–231.

LOOMIS, J. L. Communication, the development of trust and cooperative behavior. *Hum. Relat.*, 1959, **12**, 305–315.

SCHELLING, T. C. Bargaining, communication and limited war. *J. conflict Resolut.*, 1957, **1**, 19–38.

SCHELLING, T. C. The strategy of conflict: Prospectus for the reorientation of game theory. *J. conflict Resolut.*, 1958, **2**, 203–264.

SIEGEL, S., & FOURAKER, L. E. *Bargaining and group decision making*. New York: McGraw-Hill, 1960.

STONE, J. J. An experiment in bargaining games. *Econometrica*, 1958, **26**, 286–296.

ZEUTHEN, F. *Problems of monopoly and economic warfare*. London: Routledge, 1930.

(Early publication received July 20, 1960)

Journal of Abnormal and Social Psychology
1960, Vol. 61, No. 2, 190-206

VERBAL CONDITIONING OF TWO RESPONSE CLASSES: SOME METHODOLOGICAL CONSIDERATIONS[1]

JOSEPH D. MATARAZZO, GEORGE SASLOW, AND E. NELSON PAREIS[2]
University of Oregon Medical School

VERBAL conditioning has recently aroused considerable interest (Krasner, 1958; Salzinger, 1959). The research here being reported led to the investigation of several parameters of verbal conditioning which seemed to need additional study: (*a*) response class conditioned (Are *human* responses and *plural* responses equally conditionable?); (*b*) influence of different experimenters (Can one *E* cross-validate the findings of another *E*?); (*c*) influence on verbal conditioning of such subject variables as age, sex, education, intelligence (WAIS vocabulary level), Taylor anxiety level, and total time to complete the experiment; (*d*) the role of *S*'s awareness of the purpose of the experiment on his conditioning or nonconditioning performance; (*e*) the reliability of independent observer-scorers in the verbal conditioning studies here being reported; and (*f*) the influence of *S*'s base rate on his conditioning score.

METHOD

The studies were conducted on a floor housing the Division of Medical Psychology in the University of Oregon Medical School Hospital. Two soundproofed rooms, separated by a one-way mirror, and equipped with a microphone in one room and an amplifier and earphones in the observers' room were used for the experiments.

Subjects. All *S*s were paid undergraduate volunteers from Portland State College.[3] Their median age was 20, the mean was 23.5, and the range from 18 to 47. Mean educational level was sophomore year (Grade 13.4), with a range from freshman through senior year. There were 80 *S*s in all in the first four studies (20 in each), yielding a sex distribution of 48 females and 32 males. This ratio of 1.5 females to each male was roughly maintained in each of the four studies. All these *S* characteristics were obtained in the two additional studies (*N* of 20 and *N* of 10) which were conducted following the first four main studies.

Experimenters. The two male *E*s in the first four studies were a 33-year-old psychologist and a 52-year-old psychiatrist. Each *E* saw a different *S* in an odd-even design, thus yielding an original (*N* = 20) and replication (*N* = 20) experiment for the Plurals (*N* = 40) study, and an original (*N* = 20) and replication (*N* = 20) experiment for the Humans (*N* = 40) study. To further check the results of the two Plurals studies, the same two *E*s studied 20 additional *S*s (10 each, odd-even design) in a fifth study, and a third *E* (a 36-year-old male psychologist) studied 10 additional *S*s in a sixth study. All three *E*s wore their usual white hospital coats[4] during the experiment. In order to check the reliability of each *E*'s own recording and scoring of *S*'s responses during each experiment, the remaining two *E*s served as observers (*O*s) on the other side of the one-way mirror. Each *O* recorded and scored *S*'s responses independently. Thus it was possible to check *E*'s reliability of response-class judgment against each of the two *O*s and to check one *O* against the other.

Procedure. *S*s were selected from psychology classes in which it was known that neither the lectures nor the textbooks dealt with verbal conditioning. *S*s were asked to sign up "for a study on language." Each *S* was met by the designated *E* and, after being seated in the experimental room, was given these instructions:

> We are doing a study on language and how people use words. Your job today will be very simple. I want you to say words—any words at all will do, and there is no time limit—but they must be words; no sentences and no numbers. There are two people behind the one-way mirror making more elaborate counts of your words than the tally I will make. Your job, then, is simply to say words. If you have any questions about what you are to do, please ask me now because once we start I will not be able to answer your questions [pause for questions with repetition of instructions given where necessary]. All right, if you have no (additional) questions, go ahead. Say all the words you can until I tell you to stop. Go!

The conditioning procedure was as follows: In Period I (Operant or Baseline period), for *S*'s first 100 words, *E* said nothing. In Period II (Conditioning), for *S*'s next 200 words, *E* said "Good" or "That's good" every time *S* gave a word in the response class being conditioned. In Period III (Extinction), for the next 100 words, *E* again said nothing.

Because a pilot experiment had demonstrated that the first introduction of the reinforcement "Good" or "That's good," coming as it did after 100 or more words where *E* had said nothing, occasioned in some *S*s either

[1] Paper read at American Psychological Association, Cincinnati, September 1959. This research was supported by a research grant (M-1938) from the National Institute of Mental Health of the National Institutes of Health, United States Public Health Service.

[2] Now at Portland State College.

[3] We are grateful to Lester F. Beck and his staff for their generous assistance in the recruitment of these *S*s.

[4] The influence of this "status" symbol may be important in view of studies (Kanfer, 1958, pp. 443–444; Krasner, 1958, p. 152; Salzinger, 1959, p. 80; Verplanck, 1955, p. 598) suggesting a relationship between *S*'s conditionability and personal characteristics of the examiner (in this case, *E*'s implied level of prestige or social status).

startle or a response such as "Am I finished?" it was decided to introduce one such reinforcement by *E* early in the experiment. Thus, unknown to *S*, all the studies reported here used a pre-experiment period of 10 words, following the *third word* of which *E* said "Good" for every *S*, no matter what the word. Thus, *S* was introduced early to a reinforcing response by *E*. Because this introductory reinforcement was given only once, and *randomly* distributed across verbal response classes of all types, it was assumed that it would have little, if any, systematic effect on the main study. Experience demonstrated that this procedure reduced considerably, or removed completely, the previously described unwanted effects otherwise seen the first time *E* introduced the reinforcement in Period II. No rules were given *E* about the use of "Good" versus "That's good" as reinforcement. The two variations were introduced in order to make *E*'s verbal behavior sound less stereotyped. *E* was asked to use either form spontaneously and in whatever order his own spontaneity-set suggested.

Two response classes were reinforced in separate experiments: Plural Nouns and Human Responses. The former class was defined as any plural noun ending in the letter "s." This response class is relatively easy to score, although by agreement it was decided to score such words as, for example, "pants" as a *nonplural* when the context suggested a single pair. Human Responses were considered any word which clearly and unambiguously designated a person. Thus, "mother," "brother," "architect," "neighbor," etc. were included, but not "American" or "Australian," since the latter could be adjectives. Exceptions to this latter were scored when the context clearly suggested a person: e.g., "Englishman," "Frenchman," "American," "Scandinavian," "Spaniard." Such words as "necktie," "hands," etc. were not reinforced; only clear references to persons were.

Recording by both *E*s and each of the two *O*s was facilitated considerably by the use of a prepared recording sheet. This consisted of a single sheet of plain white paper on which single upright (tally) lines, in groups of five, were typed, ten groups of five or fifty to a line. The first 100, the next 200, and the last 100 were separated as Periods I, II, and III, respectively. *E* and the two *O*s drew a mark through each tally line for every word not in the response category being conditioned, and drew a circle through each tally line for every word in the appropriate category. Speed and accuracy of scoring thus were markedly enhanced.

Immediately following the conditioning procedure, each *S* was introduced to a different examiner, who individually administered the following tests in this order: (*a*) a specially prepared, open-ended, *written* questionnaire for assessing *S*'s "level of awareness" of the purpose of the experiment; (*b*) the vocabulary subtest of the Wechsler Adult Intelligence Scale (Wechsler, 1955), and the Taylor Manifest Anxiety Scale (Taylor, 1953).

RESULTS

As shown graphically in Figure 1 and in tabular form in the top half of Table 1, for the two Plurals studies, neither E_1 nor E_2 produced any significant conditioning in Period II by the use of a reinforcement following each statement by *S* of a Plural Noun in that period. With the two groups of *S*s each giving their respective *E*s approximately 16% Plural Nouns in the free or operant period, neither group increased their output of Plurals significantly with the introduction of the verbal reinforcement in Period II. As a matter of fact, E_2's group showed a slight (nonsignificant) decrease in Period II. Since no conditioning was evident in Period II, this precluded examination for extinction in Period III.

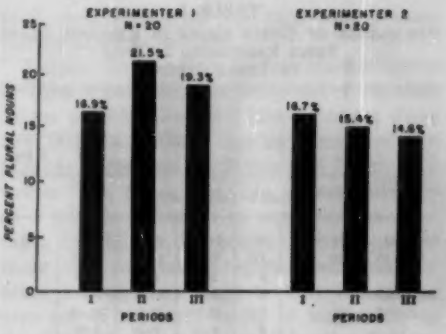

FIG. 1. Percent frequency of plural nouns in three periods for original experiment (E_1) and replication (E_2).

That neither *E* was able to condition Plural Nouns in these two studies is remarkable in view of the fact that seven out of the seven published studies of Plural Nouns reviewed by Krasner (1958, p. 160) yielded positive results. That the reinforcement used may explain our negative results is suggested by the fact that no one of these seven positive studies utilized "Good" as a reinforcement. Used instead were "mmm-hmm" (3 studies), a light (2 studies), a buzzer (1 study), and "huh-uh" (1 study). While six of these seven studies showed a significant increase in Plural Nouns during reinforcement, Greenspoon (1955) reported that the use of "huh-uh" by *E* produced a statistically significant *decrease* in *S*s' production of Plural Nouns. (Greenspoon's *S*s did show an *increase* in Plural Nouns to "mmm-hmm" as a reinforcer.) Table 1 and Figure 1 show that the trend (nonsignificant) of the data for our E_2 is in the same direction (decrease) as that of Greenspoon's, even though, following a lead from the study by Mandler and Kaplan (1956), our "Good" might be considered

TABLE 1
PERCENTAGE OF WORDS GIVEN IN RESPONSE CLASS BEING REINFORCED IN EACH OF THREE PERIODS

		Periods			
		I	II	III	*F* test
Study 1—PLURAL NOUNS					
Original Study	*M*	16.9	21.5	19.3	.86
(E_1) (N = 20)	*SD*	10.97	20.61	19.88	
	R	2-43	2-92	0-77	
Replication	*M*	16.7	15.4	14.6	.50
(E_2) (N = 20)	*SD*	14.06	10.94	10.64	
	R	1-58	0-40	1-36	
Study 2—HUMANS					
Original Study	*M*	6.7	30.9	20.7	10.60***
(E_1) (N = 20)	*SD*	4.45	23.58	28.33	
	R	1-15	6-87	3-100	
Replication	*M*	5.0	20.2	10.7	9.41***
(E_2) (N = 20)	*SD*	4.14	20.83	8.98	
	R	0-16	3-79	0-37	

*** $p < .001$.

a positive reinforcer and his "huh-uh" a negative reinforcer. This decrease for E_2 may be merely a chance occurrence. Further research will be necessary to determine the reasons for our inability to condition our *S*s on Plural Nouns in the presence of so many successful studies in the literature. It is worth mentioning, however, that colleagues from other parts of the country have reported in informal conversation similar lacks of success in conditioning Plural Nouns. Such experiences, plus our own negative results (Figure 1), would suggest that either the type of verbal reinforcement used or some other presumably identifiable variable is responsible for the different results. Examination and discussion of some of the variables which might have led to our negative results with Plural Nouns are presented later in this paper.

In contrast to the results with Plural Nouns, Table 1 and Figure 2 show that the same two *E*s were able successfully to condition the *S*s to the response class of Humans. Both their groups of 20 *S*s started off with a similar percentage (6.7 and 5.0) of Human Responses in the free-operant Period I. The mean percentage values across the three periods for both *E*s (6.7, 30.9, 20.7; and 5.0, 20.2, 10.7 for E_1 and E_2, respectively) show clear evidence of conditioning in Period II and extinction in Period III. These differences across the three periods are highly significant ($p < .001$) for each *E* by either the *F* test for correlated means (McNemar, 1955, pp. 290–296) as shown in Table 1, or (not shown) the nonparametric Friedman's two-way analysis of variance by ranks (Siegel, 1956, pp. 166–173). The latter test was employed as a check on the *F* test when a variant of Bartlett's *F* test (Edwards, 1954, pp. 327–328) on the standard deviations shown in Table 1 was significant for the Humans (but not for the Plurals) studies of both *E*s, suggesting nonhomogeneity of variance in the three periods, for this response class. (Similar nonparametric analysis of the data for the Plurals study yielded a nonsignificant *F*, again indicating no conditioning for this response class.)

Thus, as E_2 cross-validated the negative results of E_1 with Plural Nouns, he also cross-validated the positive results with Human Responses. This fact of independent cross-validation of the results by a second *E* using *S*s from the same population (see Krasner, 1958, p. 152) would suggest that the results obtained in this study are highly reliable. The additional fact (Table 1) that the base rates (obtained by the two *E*s from independent groups of 20 *S*s each) are remarkably similar for both response classes is a second indication of reliability. Such alternate-groups and alternate-*E*s

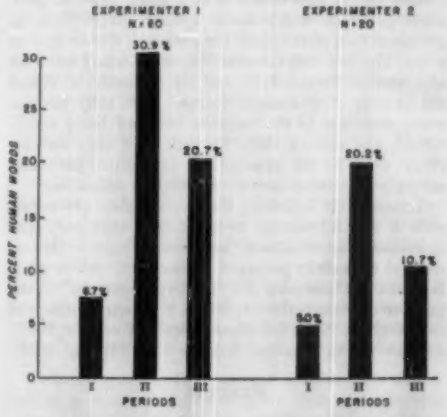

FIG. 2. Percent frequency of human words in three periods for original experiment (E_1) and replication (E_2).

TABLE 2
PERIOD COMPARISONS OF PERCENTAGE OF WORDS IN EACH OF TWO RESPONSE CLASSES

	Periods			
	I	II	III	*F*
EXPERIMENTER 1				
Plural Nouns	16.9	21.5	19.3	.86
Human	6.7	30.9	20.7	10.60***
F	14.74***	1.75	.03	
EXPERIMENTER 2				
Plural Nouns	16.7	15.4	14.6	.50
Humans	5.0	20.2	10.7	9.41***
F	12.66***	.82	.77	

*** $p < .001$.

reliability is a useful substitute for the test-retest reliability check in studies where necessary procedures (like the inquiry relating to "awareness") may make retest with the same *S*s difficult if not impossible.

Table 2 provides a statistical answer to the question: Are different response classes emitted equally frequently by *S*s in the baseline or free-operant period? It is clear from the results in this table that, while the baselines for Plural Nouns (16.9 and 16.7) and for Human Responses (6.7 and 5.0) were remarkably similar across *E*s, the percentage of Human Responses produced by the *S*s differed significantly from the percentage of Plural Nouns produced by the groups in the Plurals Study. As shown in Table 2 (reading down), analysis by conventional *F* test of the differences (16.9 versus 6.7% and 16.7 versus 5%, for E_1 and E_2, respectively) in the Period I base level production of these two response classes by the *S*s seen by each *E* (i.e., analysis across response class) was highly significant ($p < .001$). Interestingly, despite the evidence (reading across in Table 2) for conditioning in one response class and not the other, unlike the case with Period I, the differences in Periods II and III were not significant across the two response classes (reading down, see nonsignificant *F* values for Periods II and III for both *E*s in Table 2). That is, by use of reinforcement it was possible, in Period II, to bring the frequency of emission of one response class (Humans) up to the frequency of emission of the other (Plurals), even though the base rates were demonstrably different initially.

Despite the fact that the four studies shown in Table 2 used different *S*s instead of the same *S*s for assessing Period I base rates for these two response classes, the uniformity of the findings across *E*s (16.9 versus 6.7 and 16.7 versus 5.03, respectively) is further evidence of a reliable difference in response-class base rates. In addition, inspection of the means for these two response categories published by other investigators suggests that the means here are of the same order of magnitude as in other college populations (Mandler & Kaplan, 1956).

Thus, it can be concluded that Plural Nouns seem to have a lower threshold or higher emission rate than do Human Responses in the free-operant phase of verbal conditioning experiments of the type described here. If investigations in this general area continue their present proliferation, it may very well become possible to determine the free-operant frequencies of various response categories of different populations of *S*s. Wilson and Verplanck (1956) already have shown that adverbs, as a response class, have a very low operant level relative to Plural Nouns.

While significant, the results shown in Tables 1 and 2 and in Figures 1 and 2 merely represent averages or group trends. The phenomenon of "verbal conditioning" would appear to have so many theoretical as well as practical implications that it seemed wise to do a rather extensive analysis of the performance of individuals within these groups. We asked such questions as these: (*a*) Is it possible that a few *S*s in the Plural Nouns group, by starting off with large Period I base rate scores and then earning much lower Period II scores, may have influenced the results unduly by masking a majority phenomenon? (*b*) Granted that the group trends suggest that *S*'s response of Humans, as a response class, can be conditioned and extinguished by *E*'s behavior, how general is this phenomenon? Do all *S*s show it, or are the positive findings the result of a few *S*s who show extreme conditioning?

In order to examine the performance of individuals, it was necessary to define a criterion of conditioning or nonconditioning for each *S*. In the absence of any other criteria, we used

TABLE 3

OBSERVED AND EXPECTED FREQUENCIES OF INDIVIDUAL SUBJECTS FALLING IN EACH OF FOUR POSSIBLE PATTERNS ACROSS PERIODS[a]

	N	A	B	C	D	Chi square
PLURAL NOUNS						
E_1	20	5 (5)[b]	5 (5)	4 (5)	6 (5)	.20
E_2	20	4 (5)	6 (5)	4 (5)	6 (5)	.20
Both Es	40	9 (10)	11 (10)	8 (10)	12 (10)	.50
HUMANS						
E_1	20	1 (5)	18 (5)	1 (5)	0 (5)	40.20**
E_2	20	5 (5)	13 (5)	0 (5)	2 (5)	16.60**
Both Es	40	6 (10)	31 (10)	1 (10)	2 (10)	56.10**

[a] The following four patterns for the three periods were considered the most probable:

A. Base–Up–Up
B. Base–Up–Down (apparent conditioners)
C. Base–Down–Up
D. Base–Down–Down

[b] Numbers enclosed in parentheses are expected values.

** $p < .01$.

as a first step the crude index that any *S* who, relative to his own baseline, increased his production in Period II and decreased it in Period III was a "conditioner," i.e., there was a contingency between his behavior and that of the *E*. In order to account for the fact that, independent of any effect produced by *E*'s reinforcement, any given *S* could show such a pattern by chance alone, we found it necessary to define the frequency of "chance" or *apparent conditionability*. This we established as one *S* out of 4, or 25%, by the following reasoning (see bottom of Table 3). For any individual *S*, given a baseline or operant value of any magnitude (from 0 to 100%), his performance under reinforcement in Period II can be either up or down. Thus, taking Period I as a base, two directions are possible to a given individual by chance alone. With Period II as reference point, *S* also has two possibilities open to him in Period III, up or down. With such a wide possible range (from 0 to 100 in Periods I, II, and III) and our own empirical data to inspect, we decided a tie in Period II relative to Period I, or a tie in Period III relative to Period II, was so infrequent as to make these two cases not of the same order of chance magnitude as the other four cases just described. Thus, we considered the following four cases as the most probable patterns, independent of what *E* is doing: A. Base–Up–Up; B. Base–Up–Down (chance, or apparent conditioner); C. Base–Down–Up; and D. Base–Down–Down. An exception to this model would be the *S* who obtained a score of 100% in either Period I or Period II. This individual could then only go down (or tie) in the subsequent period. Since empirically no *S* showed such a pattern, we discarded consideration of this possible but unlikely exception.

If by chance alone, then, any given *S* showed our Class B, or apparent conditioning pattern, one would expect five such *S*s in each of our four studies employing 20 *S*s. Table 3 gives the obtained and expected numbers of *S* who showed each of the four patterns in the Plurals and the Humans study. The results are shown individually for each *E* and for the total group for each response class.

Inspection of the obtained and expected frequencies and of the chi square values (corrected for continuity) shows that the results yield merely chance relationships for the Plurals. For each of the two *E*s, by chance alone, 5 out of their 20 *S*s would be expected to show the B pattern. The obtained frequency of 5 and 6 for E_1 and E_2, respectively, confirms the nonsignificant group results shown in Tables 1 and 2 for this response clase. The obtained frequencies of approximately 5 for the other three patterns (A, C, and D) for each of the two *E*s also indicate chance frequencies, thus empirically confirming our earlier stated hypothesis that, in the absence of an effect due to *E*'s behavior, each of the four patterns should produce a chance frequency of 5 out of 20.

As also shown in Table 3, the analysis by individuals for the Humans study yields 18 out of 20 *S*s showing the B pattern for E_1 ($p < .01$) and 13 out of 20 ($p < .01$) for E_2 where, again, only 5 each would be expected by chance alone. The difference between the obtained frequencies (13 and 18) for the two *E*s was not significant (chi square of .80). Thus, 31 out of 40 *S*s in the two Humans studies showed conditioning. With appropriate allowance for chance (*N* of 10 out of 40), this still leaves 21 out of 40, or one *S* out of every two, who showed conditioning. The results in Table 3 for this response class clearly support the group average results shown in Tables 1 and 2 and again suggest that, with "Good" as

the reinforcement, the Humans could be conditioned (and extinguished), whereas the Plurals could not.

These results appear to offer strong support for the conclusion that certain verbal response classes are more amenable to conditioning by use of this type of experimenter verbal reinforcement ("Good") than are other response classes. In addition, the striking similarity between the individual and group results obtained by E_1 and E_2 may be considered to support the conclusion that the phenomena attendant on verbal conditioning are generalizable[6] across at least two *E*s and four different groups of *S*s.

As a further check of these two conclusions, and following upon a suggestion offered to us by Krasner and Sidowski in a personal communication (1959) that the success of our two *E*s with the Human Responses class may have been a function of more experience in verbal conditioning procedures with this class than was the case with our Plural Nouns studies, a fifth study was undertaken by our same two *E*s. In this study, 20 *S*s drawn from the same population were used. E_1 saw 10 *S*s and E_2 10 in an odd-even design. The results for these 20 *S*s (means of 13.0, 11.2, 7.9 for Periods I, II, and III, respectively) again failed to demonstrate conditioning in Period II. Furthermore, analysis by individual *S*s again failed to show ($p = .25$) a number of conditioners greater than chance expectancy.

To check on the possibility that our two *E*s were themselves somehow contributing to the lack of success with Plural Nouns, a third *E*, described earlier, was used. He had been an *O* in the previous five studies and was thoroughly familiar with the method. Using 10 additional *S*s from the same college population, he obtained Period means for Plurals of 25.8, 26.2, and 28.4, yielding no evidence of conditioning. (We have not yet investigated whether this *E*'s high base rate, 25.8, was due to the smaller *N* of 10 or whether it represents a stable *E* difference.) Thus, three *E*s and four studies of Plural Nouns failed to demonstrate conditioning for this verbal response class by the use of "Good" or "That's good" as the reinforcement. Because of our inability to condition Plural Nouns and the number of successes with this response class reported in the literature, and because we were able to condition the Human Responses class, we thought it important to investigate a number of variables which usually are not studied but which might shed light on our success with one class and failure with another.

[6] To our knowledge, only one study (Salzinger & Pisoni, 1958) has utilized a second *E* with the *same* group of *S*s as used by E_1. No study has had E_1 retest his own group of *S*s. Test-retest reliability of both types would appear to be very important to our further understanding of verbal conditioning phenomena.

SOME METHODOLOGICAL CONSIDERATIONS

Reliability of E's Scoring

There are many factors which make difficult *E*'s scoring of *S*s' responses in studies such as those reported here. Some of these are: the rapidity with which some *S*s give their serial verbal responses; the low voice level and other verbal characteristics of some *S*s, forcing *E*s to "guess" whether, for example, a word ended with "s" and was thus a plural; rapid judgments by *E* as to whether a word he is hearing for the first time falls in the response class, e.g., "astronaut" (Kanfer, 1958, pp. 443–444, makes this same point); etc. Although we did no systematic study of this variable, we were impressed in a pilot study that one of our *E*s, who "believed" in the phenomenon of verbal conditioning, tended, when in doubt about a word, to score it as "plus" when it fell in Period II, thus enhancing the chance of conditioning. Another of our three *E*s seemed always not to score these doubtful words as positive.

In view of these potential contaminating factors we used as recorders two independent *O*s in addition to *E*. This allowed us to check the reliability of each of these three people against the others. The presence of the other two recorders also reminded *E* to guard against whatever idiosyncratic sets he might have.

For each of the four studies, we were able to compute three reliability coefficients, comparing each recorder against the other two, making 12 in all. Since we computed these reliability coefficients separately for each of the three periods of the experiment, we had 36 separate Pearson *r*'s. Their range (from .96 to .99) indicated an unusually high degree of scorer reliability, thus adding additional confidence to the results here presented.

Influence of Characteristics of Individual Ss upon Conditioning

Detailed analyses were carried out on the 80 *S*s who participated in the first four studies. Inspection revealed similar results for the 30 *S*s of the last two experiments. The results for E_1 and E_2 were the same for all variables and thus will be combined.

Age. For the 40 *S*s of the Plurals study, the mean age of the 11 conditioners (22.1 years) was not statistically different from the mean age of the 29 nonconditioners (24.1 years). In the Humans group, the mean age of the 31 conditioners (23.0 years) also did not differ from the mean age of the 9 nonconditioners (25.0). Thus, with an age range from 18 to 47 for the 80 *S*s, there is no evidence that age as a variable influences verbal conditioning or nonconditioning.

Sex. The 80 *S*s included 32 males and 48 females. Of these, 15 males (47%) and 27 females (56%) showed conditioning. This slight difference between the two sexes was not significant. Verbal conditioning or nonconditioning by our two male *E*s seemed independent of the sex of *S*.

Anxiety level. For the Plurals study, the mean Taylor Manifest Anxiety Scale value for the 11 conditioners was 13.18, while for the 29 nonconditioners it was 15.31. In the Humans study, the 31 conditioners had a mean anxiety score of 11.97, while the 9 nonconditioners had a score of 13.67. For neither study, nor for both combined, were these differences significant. Thus, in each of the four separate studies, or when all 80 *S*s were combined, this research failed to show a relationship between Taylor anxiety level and verbal conditioning. This finding is in contrast to that of Taffel (1955), who did find a significant relationship between Taylor MAS score and verbal conditioning. Possibly, the answer regarding the disparate results may lie in the difference between our *S*s (college students) and his (psychotic and neurotic VA patients); or our instructions ("say words") and Taffel's (*S* was asked to make sentences using a verb and one of six pronouns); or our reinforcement ("Good") and his ("Good" for one group, a flashing light for another group, and no reinforcement for a third group); or our method of administration of the Taylor MAS (individual) and his (group).

Vocabulary level. As with the other variables, whether the comparison was of the four studies of 20 *S*s each, or for the 40 *S*s in each of the two response classes, or all 80 *S*s taken together, the mean vocabulary (standard scores) of the conditioners did not differ significantly from that of the nonconditioners. The mean vocabulary score was of the order of 13.0 for each of the four groups. Since the range of scores (8 to 19) for this variable was adequate, verbal conditioning or its absence appears to be independent of the level of *S*'s vocabulary. This may be an important finding since previous research had not dealt with what to us appeared initially to have been an important potential variable in verbal conditioning.

Total time to complete the experiment. Each *E* and the two *O*s kept a record of the total time required by *S* to complete the 400 words over Periods I, II, and III of the verbal conditioning experiment. The range for the 80 *S*s was from 11 to 81 minutes, with an average of 30.72. The mean total time of the conditioners versus nonconditioners did not differ significantly in either the Plurals or the Humans studies. Thus, neither total time nor rate (average number of verbal responses per minute) differentiated either conditioners versus nonconditioners or Plurals versus Humans.

An interesting incidental finding was the nonsignificant correlations obtained in the four studies between Taylor MAS scores and word rate for the whole experiment. The speed with which our 80 *S*s emitted their 400 words was apparently unrelated to their individual anxiety levels. This confirms the negative finding by Benton, Hartman, and Sarason (1955), who studied the word rate in a TAT situation of *S*s scoring high and low on the Taylor MAS.

In summary, with the *S*s used in this study, the results just presented for age, anxiety level, vocabulary level, and total time demonstrate no relevance of these variables to verbal conditioning. However, in each case, the analysis involved a comparison of means (by *F* test) of scores earned by two *discrete* groups (conditioners versus nonconditioners). In this analysis each *S* was weighted equally—he was either a conditioner or a nonconditioner—al-

though his own score earned individual weight within the conditioner-nonconditioner dichotomy. On the chance that such a crude, discrete dichotomization would obscure a real relationship, a score reflecting relative conditioning was obtained by simply subtracting each *S*'s Period I score from his own Period II score.

By use of both product-moment and rho correlational methods, correlations were computed between this individual conditioning score and each of these same four variables (age, anxiety level, vocabulary, and total time). In no case did the correlations reach significance, again suggesting no relationship between verbal conditioning and each of these *S* characteristics. There was a trend for total time to correlate (Pearson *r*) positively with conditioning score for both E_1 (.78) and E_2 (.36) in the Humans study, but comparable correlations were absent (.09 and −.36) in the Plurals group.

On the basis of both the analysis of variance and correlational results, and for the *S*s employed in this study, verbal conditioning, as investigated by the method here described, seems to be independent of such *S* characteristics as age, sex, anxiety level, vocabulary level, and rate of verbal responding.

When Does the Conditioning Period Begin?

Is it appropriate in verbal conditioning studies to begin recording Period II at a time or on a word number predetermined by *E*, or should the beginning of Period II be governed by *S*, individually in each case? More specifically, although *E* might decide that the first 100 words will constitute the baseline or operant period and that the next 200 words will constitute Period II or the conditioning period, what does one do in the analysis for the presence or absence of conditioning in Period II if *S* does not give his *first* response in the category being reinforced until, say, the 52nd word of the 200 words constituting Period II? Obviously, *E* cannot demand that *S* give a Human (or Plural) Response as his first word in Period II so that it can be reinforced according to schedule.

This question is not a new one for investigators in the field of operant conditioning, and they have devised a number of techniques (e.g., "shaping") to speed up the occurrence of the critical response (Skinner, 1953, pp. 91–106; Verplanck, 1955, pp. 598–599). However, since few of the many studies on verbal conditioning reviewed by Krasner (1958) and Salzinger (1959) utilized such shaping techniques, we felt that those investigators who have used the 3-period verbal conditioning methodology employed in the present study had not adequately investigated the pertinence of this question. In general, investigators have merely taken the percentage of critical words in the operant period and have put aside the question of a correction for the ordinal position of the first occurrence of a reinforced word.

Our own empirical results show that while approximately 50% of all *S*s in both the Humans and the Plurals studies gave a word in the appropriate response class—thus obtaining the reinforcement—in the first 10 words of Period II, an additional 30% in the Humans study and 15% in the Plurals study had still not given the appropriate word in the first 30 words of Period II. Even after the first 100 words of our 200-word Period II, 5% (Humans) and 10% (Plurals) of the *S*s still had not received their first reinforcement from *E*. In the Plurals study, 5% of the *S*s still had not received their first reinforcement by the 180th word.

In view of these facts, is it appropriate to begin scoring Period II on a predetermined word number? To do so tends to mask the possibility of a conditioning effect. Accordingly, we analyzed the data for Period II again, separately for E_1 and E_2, for both the Plurals and the Humans classes, this time dividing the number of words given in the response class being reinforced by the number of words (200) in that period minus the ordinal position of the first rewarded word for Period II. For example, if the first reinforcement came on the 40th word in Period II, and *S* then gave a total of 60 words for which he received reinforcement, his Period II score would be 60 divided by 160 or 37.5%. Had one not made this correction, *S*'s Period II score would have been 60 divided by 200 or 30%.

Reanalysis by this method of the data shown in Table 1 changed the mean values for Period II only slightly (21.5% to 22.3%, 15.4 to 15.9, 30.9 to 33.8, and 20.2 to 24.2). No differences across periods for the two Plurals studies emerged, and the significant ($p < .001$) dif-

ferences for the two Humans studies were maintained.

A similar question could be raised about extinction. Does it begin on a word number predetermined by *E* or on the first appropriate word for which *S* gets no reinforcement, no matter what its ordinal position? Reanalysis of the values for the Period III means shown in Table 1 again yielded corrected values only slightly different from the original ones, and the same findings were obtained: no conditioning for Plural Nouns for either *E*, and conditioning ($p < .001$) of Human Responses by both *E*s.

Verbal Conditioning and Conscious Awareness

The possibility of Thorndikian learning or, more specifically here, verbal conditioning without awareness, has received considerable attention. Excellent discussions of this general area will be found in Postman and Jarrett (1952), Adams (1957), Krasner (1958), and Salzinger (1959). Although many investigators report conditioning without awareness, Adams (1957, p. 391) reports little methodologically sound published evidence to refute the 1952 conclusion by Postman and Jarrett (1952) that "the results [of their own research] . . . show only little reliable evidence for learning without awareness" (pp. 254–255). However, Philbrick and Postman (1955), in a later modification of this earlier study, present evidence that the use of a less complex learning task "facilitated learning without awareness" (p. 424).

In view of the many published positive reports of learning and verbal conditioning without awareness contained in the reviews by Adams (1957), Krasner (1958), and Salzinger (1959), and the questions raised about these studies, especially by Adams, we decided to carry out a detailed analysis of the conscious awareness variable. Although some studies have utilized a post-experiment interview to assess *S*'s level of conscious awareness of the purpose of the experiment, we discarded this interview method after a pilot study revealed that *E*'s behavior during these interviews was too unsystematic. The interview responses of some *S*s forced *E* into a position of almost suggesting the nature of the experiment. In addition, *since the very essence of verbal conditioning experiments is that E can influence S's verbal behavior*, we felt it necessary not to use a method such that *E*'s own biases about whether or not *S*s had attained conscious awareness could in subtle ways influence *S*s' statements as to the purpose of the experiment.

Accordingly, during the post-experiment testing session, we used with each *S* a standardized procedure which employed a single sheet of white paper on which were mimeographed two open-ended statements. At the top of the page was the statement, "The purpose of the experiment was:" In the middle of the page was the statement, "My evidence for this is:" Both statements were followed by a blank space of approximately one half the page, suggesting that *S* could give an answer in considerable detail. The examiner supervised this part of the session very carefully. He took care to see that each *S* understood the two statements. He answered no questions (if any) directly or otherwise suggested answers. Later, by asking *S* (in several cases) to elaborate or otherwise rewrite an answer, he tried to minimize unscorable responses, etc.

It will be clear from the two statements used that the method employed provided information relevant primarily to whether or not *S* understood the purpose of the experiment. We did not question *S* whether or not he was aware of changes in his own verbal behavior. As Adams (1957) points out, these two questions regarding the problem of *S*'s conscious awareness (*S*'s knowing *E*'s purpose versus *S*'s knowing that his own verbal behavior had changed) are quite distinct.

Following the collection of the written conscious awareness statements from the last of the 80 *S*s of our first four studies, each *S*'s degree of "conscious awareness" was determined independently by two judges from *S*'s written responses to the two open-ended statements. The judges were given no information about *S*'s conditioning performance. A four-point rating scale (shown in Table 4) for degree of conscious awareness was developed in a pilot study and was used with the present 80 *S*s. This rating scale is presented in some detail (Columns 1 and 2) in Table 4 because, despite the importance attached to the phenomenon of conscious awareness by so many investigators, published criteria of such ratings have not been available. The illustrative ex-

TABLE 4

CRITERIA FOR LEVEL OF CONSCIOUS AWARENESS RATINGS

Awareness Level	Criteria	Examples	
		S's Statement of Purpose of Experiment	*S*'s Evidence for This Statement
1.	*S* had no idea of the purpose of the experiment, or had a completely wrong hypothesis, or made absolutely no mention of *E*'s (reinforcing) verbalizations.	"To see if I could recall past experiences."	"I realized that the things I was saying were connecting thoughts of wishes, desires, and past experiences which I had had and which surprised me somewhat when I thought how long it had been since I had thought about them."
		"I can't even begin to guess."	None
		"To gather information on whether or not the words people bring up in free-association are of real importance, or of little or no value. Also, it was supposed to reveal a person's speed of communication while under observation."	"The Doctor was keeping a count of my words, and I imagine that the number of words has something to do with an individual's ability to think—fast or slow."
2.	*S* mentioned the reinforcement, but does not connect it with a specific class of words; or *S* brings up the *possibility* of certain words being reinforced (including the correct response class), but in conjunction with other incorrect hypotheses.	"To understand the verbal contrasts between a control group and a group of patients."	"My evidence for this is: (*a*) timing by the instructor for a certain number of words; (*b*) 'good' being said by the instructor on certain words; (*c*) words expressing height, travel, and motivation received 'good'; (*d*) reused words were being counted and investigated to see how often I repeated myself."
		"An experiment on trains of thought. Also perhaps has something to do with amount of *types of words* used."	"It had something to do with trains of thought because to think of a word leads you to think of other words. Therefore everyone would have different thought patterns depending on their physical age and experience maturity. There is a chance that we were to give a definite number of words and maybe the type of word we said was important because, when I said some words he would say 'good.' This made me feel maybe I was saying what he wanted, or what was important, not just ordinary words."
		"To find out how a person relates words and by the relationship understand what areas hold their interest. This doesn't only include academic words, but social words as well. Also, I definitely think the purpose was to see how long your attention can be devoted to one particular object."	"The only evidence I have to back up my statement is when the doctor replied that my words were very good."
3.	*S* stated that *some* specific class of words is being reinforced, but names the wrong response class (e.g., *professions, my family relations*, types of *animals*, etc.); or he states the correct response class (Plurals or Humans) but does so along with a second or third (incorrect) hypothesis.	"To see if a person would continue relating words which he knew were being rewarded by the word good. Then if this reward were suddenly to stop what would be the effect upon the subjects being rewarded for like or the same words."	"Good was said after each word I said to reward that type of word [occupations which serve a purpose in society]. Also several nods were given after the word 'good' was no longer given."
		"I believe the purpose of the experiment was to have the subject name as many different occupations as possible."	"Dr. X would say, 'good,' preceding every occupation that I named, therefore I thought that that was what the purpose of the experiment was."
		"Possibly to see how large ones vocabulary is or to see how many occupational terms one knows."	"After each word such as teacher or doctor the examiner said 'good'."

TABLE 4—*Concluded*

Awareness Level	Criteria	Examples	
		S's Statement of Purpose of Experiment	*S*'s Evidence For This Statement
		"To see what thought channels a person's mind ran in and what subjects were of the most interest to him. The type of words used also helped to reveal level of education in everyday life, also what one would most generally be talking about in everyday life, if not in a test situation."	"On the small common words such as an animal, foods, clothing, etc., there was no facial change on the instructor; however, on peoples' names such as doctors, professional people, Dr. X's expression was more pleasant."
		"To see what effect certain things the interviewer did and said had on the number and kind of words I said. And also the different kinds of words and how they seemed to be selected."	"The interviewer said 'good' after certain kinds of words that I said. I immediately tried to think of more words of this same category or type. I am not sure of which kind but I'll guess: words with a great many different things or other words which would relate to them—possibly *plurals*."
4.	*S* stated that a specific response class (Humans or some very broad and similar category such as people, or Plurals) was being reinforced and *does not* state an alternate, or incorrect hypothesis.	"The experiment seemed to be a relationship of words that had to do with a particular person. The word doctor, geologist, officer, or other words that denoted a title of a person."	"Whenever I mentioned words that pertained to a person I was given a reinforcement of fine or OK from the man in charge."
		"To improve interview procedures. To be able to help lead subjects into a certain line of thought by giving a reward for correct responses."	"Whenever I mentioned words that pertained to people and their occupations I was rewarded with the word 'good.' I felt that when I received a response 'good,' that this was the line of thought I should head into."
		"To test peoples reactions to the word 'good.' The word was used in connection with a person or type of person."	"During the middle third of the test the examiner said 'good' or 'That's good' after each word that was associated with a person."
		"To determine whether the subject could recognize what words brought about a 'good' reply from the examiner, and then to see how the subject would react when the words that had once brought about 'good' replies, no longer brought about any reply at all."	"The third word I said was 'schools' which brought about a 'good' reply. When I named further words starting with an 's,' I got no reply. When I named more plural words I got no reply for some time. Then, whenever, I said a plural word, I received a 'good'. All of a sudden this stopped, but I continued to try plural words, after a few unsuccessful attempts at using singular words."

amples shown in Table 4 (Columns 3 and 4) were taken verbatim from the *S*s of the present study.

In order to check the reliability of this four-point scale, the individual ratings of the two judges were compared. The Pearson r for their independent judgments of the conscious awareness levels of the 40 *S*s in the Plurals study was .92; for the 40 *S*s of the Humans study, r was .85. The corresponding values for rho were .92 and .87. These coefficients (at the .001 level of confidence) indicated a sufficiently high degree of reliability for the rating scale shown in Table 4 to permit analysis of the relationships between each *S*'s expressed level of conscious awareness and his conditioning performance.

First to be examined were group trends for the degree to which *S*s in the two response classes were able to verbalize the purpose of the experiment. The results shown in Table 5 offer clear evidence that, as a group, the 40 *S*s in the two Humans studies were significantly ($p < .001$) more aware of the relationship between their own verbal behavior and *E*'s than were the 40 *S*s in the two Plurals studies. The difference between the awareness level associated with each of the two response classes is approximately one whole unit toward the full awareness end on the four-point conscious

TABLE 5
MEAN RATINGS OF SUBJECTS' CONSCIOUS AWARENESS LEVELS FOR TWO RESPONSE CLASSES

Judge		Plural Nouns ($N = 40$)	Humans ($N = 40$)	F
1	M	1.58	2.50	19.22***
	SD	.86	1.00	
	R	1–4	1–4	
2	M	1.55	2.25	12.10***
	SD	.86	.92	
	R	1–4	1–4	
Both	M	1.56	2.38	15.35***
	SD	.84	.92	
	R	1–4	1–4	

*** $p < .001$.

awareness scale (mean ratings of 1.56 versus 2.38 when the ratings of two judges were combined). Thus, the group data clearly suggest that the verbal conditioning obtained in the Humans studies here reported (Table 1) was associated with *S*'s greater ability to verbalize (or have conscious awareness of) the purpose of *E*'s reinforcing behavior. In many cases, the written responses clearly indicated *S*'s awareness of the contingency between *E*'s behavior and changes in his own verbal responses.

Additional evidence for this point came from a further breakdown (not shown in Table 5) of the group data. The 40 *S*s of both the Plurals and the Humans studies were again divided into conditioners versus nonconditioners, and the mean conscious awareness scores were computed separately for each response class. The results for the Humans study are that the 9 *S*s who did not show conditioning reported a significantly ($p < .01$) lower level of conscious awareness (1.56) than did the 31 *S*s who did show conditioning (2.55). For the Plurals study, while not statistically significant, the 11 conditioners showed a trend toward more awareness (1.91) than did the 29 nonconditioners (1.45). On the basis of this more stringent analysis of the group data, it can be concluded, especially in the two Humans studies, that conscious awareness as here defined was a concomitant of verbal conditioning. A similar (nonsignificant) trend held for the Plurals group data, although it should be remembered that for this response class no statistically demonstrable conditioning occurred.

In order to further evaluate the group results for these two response classes, we again investigated the performance of individuals within the groups. It was our hypothesis that if degree of conscious awareness was associated with absence or presence of verbal conditioning, then one should find more *S*s showing conditioning than not showing it as level of conscious awareness increased from Level 1 through Level 4.

The results shown in the bottom half of Figure 3 are a clear demonstration of this hypothesis for the 40 *S*s in the Humans study. Whereas one would expect by chance alone one *S* out of four to show "conditioning" at each of the four awareness levels, this chance occurrence is seen only for the 8 *S*s who reported an awareness level of 1. That is, for this stringent "no awareness" level, the 3 *S*s out of 8 (37.5%) who showed "conditioning" are not significantly different from the 2 out of 8 (25%)

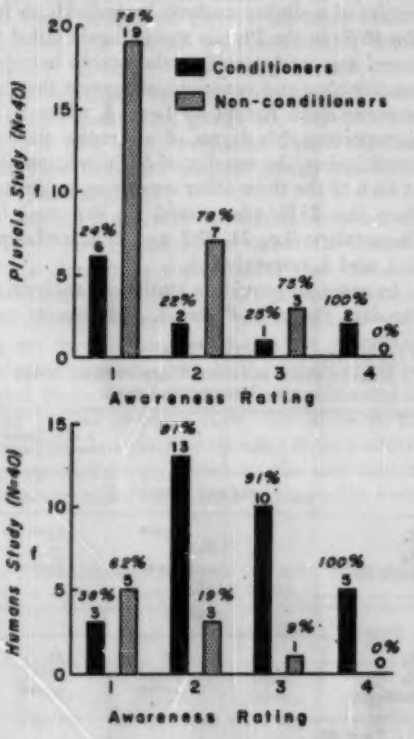

FIG. 3. Frequency distribution at each awareness level.

expected by chance alone to show an increase in Period II relative to Period I. This is not true for each of the three higher awareness levels (Levels 2, 3, and 4). For each of these three levels, the data for Humans strikingly demonstrate the fact that the number of *Ss* who conditioned far exceeded the one in four (25%) expected by chance alone (i.e., 81.2, 90.9, and 100% of the *Ss* showed conditioning at Awareness Levels 2, 3, and 4, respectively). Furthermore, despite the small *N*s involved, it is clear from these figures that, as level of conscious awareness increased from Level 2 through 4, the percentage of *Ss* showing conditioning also increased in stepwise fashion. Thus, from a chance relationship at Level 1 (37.5%), the percentage increased to 81.2% (13 out of 16) at Level 2, 90.9% (10 out of 11) at Level 3, and 100% (5 out of 5) at Level 4.

As shown in the upper half of Figure 3, the results of a similar analysis by individuals for the 40 *Ss* in the Plurals studies again failed to reveal any evidence of a relationship between conditioning and conscious awareness for this response class. Except for Level 4, where both *Ss* reporting this degree of awareness showed conditioning, the number of *Ss* "conditioning" at each of the three other levels was no greater than the 25% who would be expected by chance alone (i.e., 24, 22.2, and 25%, for Levels 1, 2, and 3, respectively).

In order to provide a statistical analysis of the data shown in Figure 3, a chi square was computed individually for each of the two response classes. This consisted of a chi square comparison of conditioners versus nonconditioners for the *Ss* with an awareness rating of 1 (minimal or no awareness) and a similar single chi square comparison for *Ss* earning all the higher awareness ratings taken as a group. For the Plurals data, the value of chi square did not begin to approach an acceptable level of significance either for the 25 *Ss* in Group 1 (chi square of .00), or for the 15 *Ss* in the combined Groups 2, 3, and 4 (chi square of .55). For the Humans data, the number of conditioners versus number of nonconditioners for the 8 minimal awareness *Ss* was equally nonsignificant (.66); whereas for the remaining 32 *Ss* it was highly significant (66.67, $p < .001$). Thus, both the analysis by groups (Table 5) and the analysis by individuals (Figure 3) demonstrate that, in the studies here reported, verbal conditioning or its absence was quite clearly associated with *S*'s degree of reported conscious awareness. It is possible, therefore, that one of the reasons for our failure to condition Plural Nouns, while other investigators have succeeded, lies in the differences in awareness levels of our *Ss* relative to the *Ss* (and other experimental conditions) employed in these other studies.

TABLE 6

RELATIONSHIP BETWEEN CONDITIONING SCORE AND AWARENESS LEVEL

	N	*r*	*r* with Transformed Scores
PLURAL NOUNS			
E_1	20	.64***	.54**
E_2	20	.31	.25
Both *Es*	40	.48***	.40***
HUMANS			
E_1	20	.38	.41
E_2	20	.49*	.55**
Both *Es*	40	.45***	.50***

* $p < .05$.
** $p < .02$.
*** $p < .01$.

The fact that the verbal conditioning of our *Ss* was correlated with *S*'s level of conscious awareness was demonstrated in still another way. As previously mentioned, a dichotomy into conditioner versus nonconditioner, such as is presented in Figure 3, gives each S who shows conditioning equal weight to every other *S* who shows conditioning, regardless of magnitude. In order to allow for each *S*'s actual conditioning performance, the previously described conditioning score for each *S* (the difference between *S*'s Period II minus Period I score) was correlated with his reported conscious awareness score. In view of the high correlations between the two judges for this awareness score, the two judgments were pooled for an average awareness score.

The results of this analysis are shown in Table 6. Due to the restricted range (1 to 4) of the awareness ratings, as well as the presence of skewed distributions of scores for both variables, a simple square root transformation (Snedecor, 1946, pp. 445–447) was made. These

results are shown along with the Pearson r values in Table 6. Values for rho also were computed and yield similar results.

The data presented show a moderate but significant correlation between magnitude of conditioning score (increase from Period I to Period II) and level of later reported conscious awareness. The positive correlations and an inspection of the frequency distribution of conditioning scores indicate that the greater the increase from Period I to Period II, the higher was *S*'s later reported level of conscious awareness. Three of the 40 *S*s in the Human Responses class showed a decrease in Period II relative to Period I, whereas 20 of the 40 *S*s in the Plurals study showed such a decrease. Fourteen of these 20 *S*s fell in awareness Level 1. For Levels 2, 3, and 4, the frequency distribution for Plurals (i.e., the conditioning score), like that for Humans across Levels 1, 2, 3, and 4, showed a positive correlation between amount of conditioning and degree of later reported conscious awareness. This finding of a significant positive correlation between magnitude of conditioning score and level of awareness for the Human Responses is not surprising in view of the results shown in Figure 3. That the dichotomous analysis shown in Figure 3, as well as the chi square analysis, did not demonstrate such a similar relationship for the Plural Nouns class, whereas the correlational analysis shown in Table 6 for this same response class did, would suggest that magnitude as a measure of conditioning related to level of conscious awareness may be more sensitive than dichotomous classification. In any event, all analyses demonstrate the same finding: for our *S*s, verbal conditioning occurred in those *S*s who later could verbalize the purpose of the experiment. This was especially true for the Human Responses class (Figure 3 and Tables 5 and 6), but also true for one analysis of the Plural Nouns data (Table 6). The minor differences shown in Table 6 between the results obtained by the two *E*s appear to be nonsystematic and, thus, probably are due to random sampling errors.

Influence of S's Base Rate on Verbal Conditioning

In view of the significant differences in the base rates of the Humans versus Plurals response classes (Table 2), it appeared wise to investigate further the role of these base rates in conditioning, especially as base rates and conditioning might in turn be related to *S*'s level of conscious awareness. Salzinger and Pisoni (1958, p. 88) reported a finding which bears indirectly on this question. They reported a cross-validated rank-order correlation of .46 between the number of interview Affect responses given by 20 schizophrenic *S*s in a 10-minute base period and the number of Affect responses given in the next 10-minute (conditioning) period. Sixteen additional *S*s, serving as a control group, and with whom *E* attempted no conditioning, showed a still higher correlation (rho of .85, $p < .01$) between the first and second 10-minute periods. Thus, although smaller in magnitude (rho of .46), there still appeared to be a correlation between base period and conditioning period levels for this (Affect) response class. However, it also is clear from these rho values that when *E* applied his verbal reinforcement ("mmm-hmm," "I see," "yeah," and other forms of agreement) and thereby produced conditioning in the experimental group, the correlation between base rate level and Period II level dropped from .85 (control group) to .46 (experimental group).

Similar rank-order correlations between Period I and Period II are shown in Table 7. It is clear from these results that, for both response classes, there is a significant positive correlation between the percentage of words given in the base period and the percentage of similar words given in the conditioning period. Since we failed to obtain conditioning for the Plurals response class, the 40 *S*s in this response class may be thought of as a control group crudely analogous to the one used by Salzinger and Pisoni. Such a conception would

TABLE 7
RELATIONSHIP BETWEEN SUBJECTS' PERIOD II AND BASE RATES

	N	Plural Nouns rho	Humans rho
E_1	20	.68**	.20
E_2	20	.56**	.47*
Both *E*s	40	.61**	.36*

* $p < .05$.
** $p < .01$.

offer a tentative hypothesis for explaining the somewhat smaller, albeit still significant, rhos obtained by us for the Human Responses class. That is, it is possible to conclude that the correlation between *S*'s verbal behavior in Period I and Period II may be positive but small when *S* is aware of the contingency between his own behavior and *E*'s (our Humans study), but may be much greater either when *E* is applying no reinforcement (Salzinger and Pisoni), or when *S* is unaware of the reinforcement (our Plurals study), or when reinforcement is having no effect for other reasons.

In view of this demonstrated small positive correlation between base rate and Period II response level, especially in the Humans study, we decided next to investigate the relationship between base rate and the earlier described magnitude of conditioning score, i.e., how much an *S* increased (or decreased) his score from Period I to Period II. We reasoned that this conditioning score might turn out to be independent of *S*'s base level because it represents the *difference* between Period I and Period II scores.

Because of the presence of skewed distributions, both rhos and Pearson *r*'s were used for computing the correlation between each *S*'s base rate and his subsequent conditioning (difference) score. For the Human Responses class, and for each *E* and for both *E*s combined, the two types of correlations between base rate and conditioning score were found to be near the zero level. Thus, conditioning of the Human Responses class, as demonstrated in Table 1 and in the analysis by individuals (Table 3), appears independent of each *S*'s own operant level.

Interestingly, the same lack of relationship was not true for the Plurals response class. For this category, a small negative correlation ($-.37$, $p < .02$) was obtained between base rate for Plurals and conditioning score. The *S*s of both *E*s obtained these small negative correlations, indicating that the higher the base rate, the smaller was the increase from Period I to Period II and vice versa.

One possible explanation for this finding of a small negative correlation in the Plurals studies but not the Humans studies is that the negative correlation between base rate and conditioning score is unrelated to the magnitude of the conditioning score but, rather, is related to the magnitude of the base rate. This hypothesis is based on the demonstrated ($p < .001$) difference in base rates for the two response classes (Table 2)—16% for Plurals versus 6% for Humans—and the correlated fact that an *S* with a high base rate (as in the two Plurals studies) did not have the same opportunity to increase his score in Period II to the same degree as an *S* with a lower base rate (as in the two Humans studies). Thus, the *S* with a large base rate score would have a higher probability of earning a lower conditioning score, thus yielding a negative correlation between base rate and conditioning score, than would the *S* with a smaller base rate score. Until further research bearing on this hypothesis is completed, we believe it to be an adequate one to explain our findings.

Thus the results of our own study (Table 7) and Salzinger and Pisoni's (1958) suggest a significant positive correlation between actual scores in Period I and actual scores in Period II. The small significant negative correlation (between Period I score and conditioning score) just described indicates a qualification to this inference. Whereas an *S*'s Period II score can be predicted, within limits, from his own Period I or base level score (Table 7), the magnitude of his change from Period I to Period II is inversely related to the magnitude of his base level score for one response class (Plurals) and has no relationship for another response class (Humans), presumably because this latter response class has a lower operant or base rate. Furthermore, the negative correlations obtained for one response class (Plurals) and not for another (Humans) again demonstrate the necessity for careful study of different verbal response classes (including their free-operant levels) before the results of verbal conditioning studies are generalized too far.

Why Are Some Verbal Response Classes Easier to Condition than Are Others?

At this stage of our knowledge no adequate data exist to answer this question of differential effects over response classes. Although the results of our own studies show that *S*'s degree of conscious awareness is correlated with whether or not he shows conditioning, the fact still remains that we were able to produce both this higher level of awareness and a change in *S*'s

behavior in Period II ("conditioning") with the Human Responses category but not with the Plural Nouns category. Thus, the two response classes must differ along at least one crucial dimension. Table 2 shows that the free-operant frequency of emission of these two classes of responses differs significantly. It is our own belief that the two response classes differ in still another way, in their level of difficulty as concepts. That is, considered purely in terms of difficulty of concept formation, it is reasonable to predict that *S*s could more readily abstract or perceive the similarity among five verbal responses like mother, neighbor, friend, architect, plumber than among books, apples, cars, shoes, flowers. If this is true, then Humans as a response class may be more easily conditioned than the more difficult concept Plurals. This would be especially true if *S*'s level of conscious awareness is related, as is the case here, to his conditioning performance. In any event, the difficulty level of the concept being conditioned appears to be a variable worth exploring. This variable may turn out to be as important as the others already identified (reinforcement used, base rates, *S*'s level of reported conscious awareness, etc.).

Another possible explanation for our success in conditioning Humans and not Plurals has been offered by B. F. Skinner (personal correspondence, 1960). It involves his concepts of the discriminative stimulus, S^D, and the prior reinforcement and conditioning history of the individual. His suggestion runs as follows:

> It seems to me quite reasonable that the so-called reinforcement in period two acts more as a discriminative stimulus than as a reinforcement. In talking to the lady on my right at a dinner party, I may try various subjects without much success. When I hit on something which gets a nod of interest, I continue along the same line, not so much because the response has been reinforced, as because the response plus the reinforcement act as an S^D indicating that this is the kind of audience which reinforces certain kinds of responses. The class of responses has already been learned in the past. The first "Good" in period two may be closer to a green light which serves as a discriminative stimulus for "more of the same." If this is the case, you could probably show that the class of *humans* has already been brought under audience control where the class of *plurals* has not. Thus, one could discover that a given listener is interested in *people*, although it would be quite rare to discover that a given listener is interested in *plurals*. This would throw the explanation back on the prior conditioning of the subject with respect to audience variables . . . .

With notable exceptions (Salzinger, 1959, pp. 85–87), investigators working in the area of verbal conditioning have tended to ascribe the conditioning effect to the reinforcement used by *E*, rather than to think in terms of S^D and the individual's prior differential conditioning history to disparate verbal response classes.

Another variable which may have been responsible for our success with the Human Responses class and failure with Plurals may be an "expectancy-set" induced in our *S*s by the physical setting in which the experiments were conducted. Since the *S*s of this study were studied in the laboratories of the psychology-psychiatry service of a large medical center, the *S*s, in line with cultural conditioning regarding psychiatry, may have been "set" to talk about themselves and other people, thus showing a conditioning effect to this "expected" Human Responses class and not to the "not-expected" Plurals class. Thus "situation-induced expectancy" as a set-variable may be pertinent in understanding our results. This is especially so since our *E*s wore their usual white hospital coats during the conditioning experiments, and thus may have enhanced the set of seeking responses of an interpersonal (Humans) nature in our *S*s.

SUMMARY

Two *E*s, in replication experiments, independently studied two groups of 20 college undergraduates each, making a total of 80 *S*s in all. As additional checks on the results, the same two *E*s studied a fifth group of 20 *S*s, and a third *E* studied a sixth group of 10 *S*s. Verbal conditioning was attempted in separate studies with two verbal response classes: Plural Nouns and Human Responses. Directions called essentially for *S* to "say words, but not sentences or numbers." The *E* used three periods of 100, 200, and 100 words, constituting the operant, conditioning, and extinction periods, respectively. The statement "Good" or its variant, "That's good," was used by *E* as the reinforcement in Period II.

Results indicate that neither *E* could condition Plural Nouns, whereas both *E*s successfully conditioned the Human Responses class. Thus, response class is an important variable

in verbal conditioning. Since a number of different analyses all revealed that for both *E*s the successful conditioning of the Human Responses class was associated by *S*s in these groups with a higher level of conscious awareness of the purpose of the experiment than was the case with *S*s in the Plurals groups, it was demonstrated that the presence of "conscious awareness" was a concomitant of conditioning. Level of *S*s' conscious awareness was determined independently by two judges from written statements to two open-ended questions obtained from *S* immediately after the verbal conditioning experience. The four-point rating scale used for these judgments was shown to have high interjudge reliability. Further analysis revealed that under the conditions of these studies, verbal conditioning was independent of such characteristics as age, sex, vocabulary level, (Taylor) anxiety level, and word rate.

In order better to understand the phenomenon of verbal conditioning, a number of additional methodological considerations were discussed: reliability of *E*'s scoring, determination of when the conditioning period begins, influence of *S*'s base rate on verbal conditioning, and influence on verbal conditioning of the level of difficulty of different response classes as concepts. Analyses of the data of the present studies were used in the presentation of these methodological considerations.

REFERENCES

ADAMS, J. K. Laboratory studies of behavior without awareness. *Psychol. Bull.*, 1957, **54**, 383–405.

BENTON, A. L., HARTMAN, C. H., & SARASON, I. G. Some relations between speech behavior and anxiety level. *J. abnorm. soc. Psychol.*, 1955, **51**, 295–297.

EDWARDS, A. L. *Statistical methods for the behavioral sciences*. New York: Rinehart, 1954.

GREENSPOON, J. The reinforcing effect of two spoken sounds on the frequency of two responses. *Amer. J. Psychol.*, 1955, **68**, 409–416.

KANFER, F. H. Verbal conditioning: Reinforcement schedules and experimenter influence. *Psychol. Rep.*, 1958, **4**, 443–452.

KRASNER, L. Studies of the conditioning of verbal behavior. *Psychol. Bull.*, 1958, **55**, 148–170.

MCNEMAR, Q. *Psychological statistics*. New York: Wiley, 1955.

MANDLER, G., & KAPLAN, W. K. Subjective evaluation and reinforcing effect of a verbal stimulus. *Science*, 1956, **124**, 582–583.

PHILBRICK, E. B., & POSTMAN, L. A further analysis of "learning without awareness." *Amer. J. Psychol.*, 1955, **68**, 417–424.

POSTMAN, L., & JARRETT, R. F. An experimental analysis of "learning without awareness." *Amer. J. Psychol.*, 1952, **65**, 244–255.

SALZINGER, K. Experimental manipulation of verbal behavior: A review. *J. gen. Psychol.*, 1959, **61**, 65–94.

SALZINGER, K., & PISONI, STEPHANIE. Reinforcement of affect responses of schizophrenics during the clinical interview. *J. abnorm. soc. Psychol.*, 1958, **57**, 84–90.

SIEGEL, S. *Nonparametric statistics for the behavioral sciences*. New York: McGraw-Hill, 1956.

SKINNER, B. F. *Science and human behavior*. New York: Macmillan, 1953.

SNEDECOR, G. W. *Statistical methods*. Ames, Iowa: Iowa State College Press, 1946.

TAFFEL, C. Anxiety and the conditioning of verbal behavior. *J. abnorm. soc. Psychol.*, 1955, **51**, 496–501.

TAYLOR, JANET A. A personality scale of manifest anxiety. *J. abnorm. soc. Psychol.*, 1953, **48**, 285–290.

VERPLANCK, W. S. The operant, from rat to man: An introduction to some recent experiments on human behavior. *Trans. NY Acad. Sci.*, 1955, **17**, 594–601.

WECHSLER, D. *Manual for the Wechsler Adult Intelligence Scale*. New York: Psychol. Corp., 1955.

WILSON, W. C., & VERPLANCK, W. S. Some observations on the reinforcement of verbal operants. *Amer. J. Psychol.*, 1956, **69**, 448–451.

(Early publication received June 16, 1960)

Journal of Abnormal and Social Psychology
1960, Vol. 61, No. 2, 207-215

ANXIETY IN ACADEMIC ACHIEVEMENT SITUATIONS[1]

RICHARD ALPERT
Harvard University

AND

RALPH NORMAN HABER
Yale University

As part of the effort to define and measure the variables of which test performance is a function, this study was designed to evaluate, both experimentally and theoretically, the paper-and-pencil instruments currently being used in American research for the measurement of individual differences in anxiety as it affects academic achievement performance. In addition, it includes a description of a new achievement-anxiety scale which has been devised to indicate not only the presence or absence of anxiety, but also whether the anxiety facilitates or debilitates test performance.

Three separate problems are considered: (*a*) the relationship between scales which are designed to measure general anxiety and scales specifically designed to measure test anxiety (specific anxiety scales), and a comparison of the relative efficacy of the general and specific scales as predictors of academic achievement performance; (b) the relationship between the construct of anxiety and that of aptitude and the methodological problems involved in separating these two operationally; and (*c*) the direction of the effect of anxiety upon academic achievement performance.

Procedure and Subjects

Procedurally, this research involved the administration of a variety of anxiety scales to introductory psychology students and to freshmen at Stanford University. The scales used included the Taylor Manifest Anxiety Scale (MAS) (Taylor, 1953), the Welsh Anxiety Index (AI) (Welsh, 1952), the Freeman Anxiety Scale (AS) (Freeman, 1953), the Mandler-Sarason Test Anxiety Scale (TAS) (Mandler & Sarason, 1952), and the Achievement Anxiety Test (AAT), which has two scales discussed below. Analysis of the data involved intercorrelating these scales and studying their relationship to a measure of verbal aptitude, the Scholastic Aptitude Test (SAT) (College Entrance Examination Board, 1956), and to a set of academic performance indices, including college grade-point average and the final examination, mid-term examination, and course grades in the introductory psychology course. Only male *S*s were included in the data analyses. For the students enrolled, the scales were either administered during the first class meeting of the quarter to the entire class (Fall 1955, Winter 1956, and Spring, 1956, with *N*s of 93, 92, and 98, respectively), or as part of smaller group experiments which were required of all students (Spring, 1956, subsample, $N = 40$). Therefore, these data are representative of the population of the students from the introductory course in psychology, which is taken by 75% of the students in the university. For the freshmen, the scales were group administered as part of the freshman evaluation program.

The scores for the Scholastic Aptitude Test (administered to the students during their senior year in high school) and the grade-point averages were obtained from the Registrar of the university. The final examination, course, and midterm grades were obtained from class records. All of these data were collected without the *S*s' awareness either that it had been collected or that data of this nature would be related in any way to their responses on the anxiety scales.

In view of the finding of Davids (1955) that scores on anxiety scales are affected by the *S*s' particular motivation for filling out anxiety questionnaires, it must be noted that *S*s completed the present questionnaires with the understanding that the data were being collected as part of a large ongoing research program conducted by the Psychology Department and that their answers would be used for research purposes only. In some instances, *S*s' affiliation motivation might have been aroused when, for example, the entire class was asked to fill out a questionnaire as a favor to the experimenter; at times when the questionnaire was administered in experimental session, the fulfillment of course requirements was the major motivation. In all situations, however, it was emphasized that the data would *not* be made available either to the university administration or to the faculty and that no selection of any kind would be based on the results of the questionnaire. There is no reason to believe that motivational differences in any way differentially affected the responses.

General versus Specific Anxiety Scales

The first concern in constructing a measuring instrument which will predict the effect of anxiety in academic performance is one of defining the population of experiences and behaviors which the item responses are to reveal. There are current in the literature two

[1] The support provided by the National Science Foundation (G-1787 and G-3045) is gratefully acknowledged. We also wish to thank R. R. Sears, Q. McNemar, and Jan Pierce for their help and support, and the Stanford University Office of the Registrar and Computing Center for their cooperation.

Part of this report appeared as a portion of a doctoral dissertation at Stanford University by the first author.

divergent positions regarding this matter. One position is exemplified by Taylor (1953) with her Manifest Anxiety Scale and the other by Mandler and Sarason (1952) with their Test Anxiety Scale.

Taylor's general anxiety scale is made up of items drawn from the MMPI and is concerned with a wide variety of situations other than test taking. Underlying the construction of the MAS is a theoretical assumption that there is a relatively constant "level of internal anxiety or emotionality," and also "that the intensity of this anxiety could be ascertained by a paper-and-pencil test consisting of items describing what have been called overt or manifest symptoms of this state" (Taylor, 1953, p. 285). If Taylor is correct in positing such a general anxiety state, then it should follow that a single measure of a set of manifest anxiety responses gathered from many situations would be an adequate predictor of the presence and effects of anxiety responses in *any* situation, whether it be eyelid conditioning (Spence & Taylor, 1951; Taylor, 1951), therapeutic sessions (Peck, 1951), serial learning (Taylor & Spence, 1952), or academic performance. At present, however, there is some evidence (Child, 1954; Mandler, 1954) of *intra*-individual differences in anxiety both in content and intensity from one situation to another. This evidence warrants further examination of the basic general-anxiety-state assumption underlying the construction of the Taylor MAS-type scale before applying such a scale as extensively as has been done in the past. Too often, conclusions have been drawn on the basis of correlations with the MAS regarding the presence or absence of anxiety or the effects attributable to anxiety without due consideration of the possible limitations of a general scale of this type as a sensitive indicator of anxiety in any limited, recurring type of situation.

Mandler and Sarason, who represent the alternative current position, maintain that the items composing the measuring instrument should be concerned with the specific situations in which it is to be used. These investigators, studying stress in academic achievement situations, devised the Test Anxiety Scale (TAS), "...a questionnaire which was specifically concerned with the *S*s attitudes and experiences in a testing situation" (Mandler & Sarason, 1952, p. 166). Thus, the TAS provides a score indicating the recalled intensity of certain experiences and behaviors immediately antecedent to or concomitant with the taking of various types of examinations in the past. The implication of this type of scale is that the increased situational specificity of its item content will allow for a more sensitive measurement of anxiety and its effect in the academic achievement situation. It follows that for other situations other specific scales will be needed; and, only if these specific scales, each shown to be valid in its own situation, turn out to be highly intercorrelated may one posit a general anxiety state and thus justify a single measure of anxiety for all situations.

The relative merits of situational specificity versus generality of item content can be evaluated by two statistical techniques. The first is a comparison of the intercorrelations among a variety of general and specific scales administered to the same subjects. The Taylor MAS, the Welsh AI, and the Freeman AS are used as measures of general or manifest anxiety; the Mandler-Sarason TAS and the two scales of the AAT are used as measures of specific anxiety. High intercorrelations among all six scales would support the theoretical assumption of a single underlying state, and thus the appropriateness of a single general measure. On the other hand, low intercorrelations would support the use of scales which are specific to the situations in which they are to predict. Table 1 reports the intercorrelations of the measures of anxiety for an N of 40 (the only sample that received all six scales). Because in all instances inspection of the scatter diagrams revealed no departure from linearity, only product-moment correlation coefficients were computed.

As Table 1 indicates, the correlations among the general anxiety scales range from .32 to .39. The correlations between the general and specific scales range from .24 to .38. The correlations among the specific anxiety scales range from .40 to .64, all of which appear to be higher than any of the correlations involving the general scales. This observation is supported by significance tests of the differences between correlations, which indicate that in most instances the intercorrelations among

TABLE 1
INTERCORRELATION AMONG SIX MEASURES OF ANXIETY
(N = 40)

Test	REL	AI	AS	TAS	AAT−	AAT+
General Anxiety Scales:						
MAS	.89	.39**	.32*	.32*	.38*	−.33*
AI	.84		.34*	.28	.37*	−.25
AS	.73			.38*	.30	−.24
Specific Anxiety Scales:						
TAS	.82				.64**	−.40*
AAT−	.87					−.48**
AAT+	.83					

Note.—Reliabilities of the scales appear in Column 1.
* Based on the nonoverlapping items. When overlapping items are included, r = .39.
* Significant at .05 level.
** Significant at .01 level.

TABLE 2
CORRELATIONS BETWEEN SIX ANXIETY SCALES AND FOUR MEASURES OF ACADEMIC PERFORMANCE
(N = 40 to 379)

Test	Grade-Point Average	Course Grade	Final Exam Grade	Midterm Exam Grade
		Psych 1	Psych 1	Psych 1
General Anxiety Scales:				
MAS	.01	−.08	−.02	−.19
AI	−.04	−.05	−.03	−.22*
AS	−.06	.14	.15	—
Specific Anxiety Scales:				
TAS	−.24*	−.21	−.16	−.32*
AAT−	−.35*	−.26*	−.28*	−.25*
AAT+	.37*	.23*	.26*	.21

* Significant at .05 level.

the specific scales are significantly higher than the correlations involving the general scales either with each other or with the specific scales. This seems to throw some doubt on the comparability and, therefore, substitutability of a general anxiety scale for a specific anxiety scale.

A factor analysis was also performed (Alpert, 1957) on this matrix of correlations, but although the above conclusions were supported by this analysis, such support is little more than suggestive because of the hazards of drawing conclusions from a factor analysis based on only 40 cases.

A second technique for determining the substitutability of the general for the situationally specific anxiety scales, far more rigorous than merely intercorrelating the scales, is to compare the two types of scales as predictors of various measures of academic performance. If the general scales predict as well as the specific scales, there is justification for the use of a single general measure for the detection of anxiety in all situations. If, on the other hand, the specific scales are able to account for significantly more of the variance in the performance measures than are the general scales, then there is justification for the use of scales which are specific to the situation in which they are used.

Table 2 presents the correlations between the six anxiety scales and four measures of academic performance. These are average correlations, based on combined samples with total Ns ranging from 40 to 379. Thus, these relationships are based on all samples for which both the appropriate anxiety and performance scores were available. Before averaging was attempted, a check was made to be sure that neither the means and the standard deviations of all the samples nor the correlations to be averaged were significantly different. No differences were found in any of these comparisons.

The data summarized in Table 2 indicate that the specific anxiety scales are more often significantly correlated with academic performance measures than are the general anxiety scales. Accepting the 5% level of significance (two-tailed) as a criterion, only one of the 11 available correlations for the general scales reached significance, while 9 of the 12 correlations involving the specific scales reached significance.

The implications of the findings are reasonably clear. Specific anxiety scales and general anxiety scales measure, to a significant extent, something different. Furthermore, it appears that the variable which the specific scales measure, and which the general scales do not, is involved in academic performance to such an extent that the specific scales are better predictors of academic performance than are the general anxiety scales.

One may tentatively conclude, therefore, that further use of a general anxiety scale as

TABLE 3
CORRELATIONS BETWEEN SIX ANXIETY SCALES AND A MEASURE OF VERBAL APTITUDE (SAT)
(N = 40 to 379)

Test	Verbal Aptitude		
	N	r	p
General Anxiety Scales:			
MAS	198	+.10	*ns*
AI	153	+.13	*ns*
AS	40	−.24	*ns*
Specific Anxiety Scales:			
TAS	154	−.18	.05
AAT−	379	−.29	.001
AAT+	379	+.21	.001

an appropriate operation for the measurement of academic achievement anxiety is unwarranted. However, the next section, on the relationship of these various anxiety scales to aptitude must be considered before this can become a final conclusion.

THE RELATIONSHIP BETWEEN ANXIETY AND APTITUDE

In evaluating the efficacy of the various scales in the academic situation, one must take cognizance of the relationship between the anxiety measures and intellectual ability because interest is centered in a scale which predicts performance variance attributable to something other than aptitude. On this basis, it seems desirable for an anxiety scale to correlate highly with performance and only slightly with aptitude.

The equivocal results reported in the literature regarding the relationship between anxiety and aptitude become more clear if one takes into consideration the type of anxiety scale used, the heterogeneity of the population measured, and the nature of the aptitude measure (most importantly, whether it was timed and, therefore, under pressure or untimed). (See Alpert, 1957, for a more extensive review of these variables.) Since the present study is concerned with a highly intellectually homogeneous college population and with an aptitude test given under pressure conditions, the literature involving comparable populations and test conditions is of most direct concern. It indicates that the general anxiety scales are not related to timed aptitude tests in homogeneous college populations (Davids & Eriksen, 1955; Schulz & Calvin, 1955), whereas specific anxiety scales are related to this kind of aptitude test (Sarason & Mandler, 1952). The data presented in Table 3 support these findings. Here the general anxiety scales are not significantly related to aptitude, but all the specific scales are significant beyond at least the 5% level.

There are four possible explanations for this difference:

1. The specific scales are in part a measure of intellectual ability. That is, for some reason, intelligence, *independent of actual anxiety level*, affects the individual's response to a specific anxiety measure. General scales are free of an intellectual ability component according to this argument.

2. A specific anxiety scale is an appropriate vehicle for rationalization or justification of poor academic performance. Because people of lower intelligence are more apt to have experienced the effects of the results of poor academic performance in the past, more of them would be inclined to use such rationalization than would people of higher intelligence, and the specific scales are more sensitive to rationalization attempts.

3. There are stress cues connected with the taking of timed college aptitude tests. These cues elicit anxiety responses which affect the aptitude test performance. The specific scales are sensitive to the presence of anxiety of the type reflected in the aptitude tests, but the general scales are not similarily sensitive.

4. The more intelligent the individual, the less anxiety he manifests in test situations because he has less objective reason to fear the experience. The specific scales, which are able to measure this interaction, are, therefore, a sensitive index of anxiety in a college population, whereas the general scales are not.

The last of these arguments is the only one which maintains that the obtained correlation is a measure of the relationship between the two constructs of anxiety and aptitude. The first three suggest that the correlation between the measures is an artifact which appears because of poor test construction or the conditions of measurement. While the present studies were not designed to evaluate

these various explanations, they do provide suggestive evidence of their relative validity.

The first explanation holds that if an anxiety scale which is correlated with an aptitude measure is, because of some characteristic such as its format, simply a measure of aptitude itself, then one would expect that the aptitude and anxiety measures would account for the same variance in academic performance. If this were, in fact, the case, then a multiple correlation based on the best weighted combination of aptitude and anxiety would not be expected to account for significantly more of the variance in academic performance than would aptitude alone. Tables 4 and 5 present multiple correlations of aptitude and anxiety, predicting grade-point average and final examination grade. Also, they indicate the significance of the addition of the anxiety measure in the predictions. The general anxiety scales are not included in the tables because they are not correlated with achievement performance and hence, could make no contribution to the prediction of such performance.

The debilitating scale of the AAT adds significantly, over and above aptitude, to the prediction of the grade-point average (GPA) in two out of three instances, and in one out of three for the final examination grade. The facilitating scale of the AAT adds significantly, over and above aptitude, to the prediction of GPA in the three instances reported, and in two out of the three instances for the final examination grade. The Mandler-Sarason TAS adds significantly to prediction of final examination grade, but not for GPA, in the only samples which data are available. These findings refute the argument that the specific scales which are correlated with aptitude are simply measures of aptitude, because the results indicate that specific anxiety scales are able to account for variance in academic performance other than that accounted for by a measure of aptitude. This does not mean to say that these anxiety instruments are not, in part, measures of intellectual ability; it does say, however, that they are something more than that—and that therefore they are able to account for added variance in academic performance, and, as such, are valuable supplements to a measure of aptitude.

TABLE 4

MULTIPLE CORRELATIONS PREDICTING GRADE-POINT AVERAGE (GPA) FROM A COMBINATION OF VERBAL APTITUDE AND ANXIETY

Test	N	R^a	r^b	r^c	p^d
AAT−	93	.46	−.45	.27	.001
AAT−	92	.29	−.08	.29	*ns*
AAT−	98	.53	−.40	.43	.001
TAS	40	.46	−.27	.41	*ns*
AAT+	93	.42	.36	.27	.01
AAT+	92	.43	.32	.29	.01
AAT+	98	.58	.50	.43	.001

[a] GPA · Aptitude & Anxiety.
[b] GPA · Anxiety.
[c] GPA · Aptitude.
[d] For adding anxiety to aptitude.

TABLE 5

MULTIPLE CORRELATIONS PREDICTING FINAL EXAMINATION GRADE FROM A COMBINATION OF APTITUDE AND ANXIETY

Test	N	R^a	r^b	r^c	p
AAT−	93	.49	−.48	.30	.001
AAT−	92	.42	−.02	.42	*ns*
AAT−	98	.35	−.34	.16	*ns*
TAS	40	.40	−.35	.25	.05
AAT+	93	.36	.25	.30	.05
AAT+	92	.47	.23	.42	.05
AAT+	98	.32	.28	.16	*ns*

[a] Final examination grade—Anxiety & Aptitude.
[b] Final examination grade—Anxiety.
[c] Final examination grade—Aptitude.

The second explanation holds that the significant correlation of specific anxiety scales with aptitude is really the result of a mediating process of rationalization and justification for poor performance. In order to evaluate this argument, the entering freshman class was given a specific anxiety scale. Freshmen entering Stanford have, in the majority, a past history of similar academic success experiences in high school, and they have had no past experience with the college academic performance criteria which anxiety scales were to predict. The AAT was administered to the entire freshman class at Stanford during their orientation week in the Fall of 1955. After the end of their first year, a random sample of 96 was drawn and the AAT was correlated with their grade-point averages. Correlations of .42 with the facilitating scale and −.29 with the debilitating scale were obtained. Both correlations are significant beyond the 1% level. Comparing these correlations with those obtained for a sample of 70 upperclassmen

for whom exactly comparable data were on hand, no significant differences were found.

This is a weak refutation of the second explanation, at least at the college level. But, of course, college students have had anxiety experiences regarding high school achievement, and thus, it is still possible that the freshmen could be rationalizing via the AAT for their high school grades. But the two scales were correlated with high school grade-point average .29 and −.21, respectively. These correlations are both significant at the 5% level, and they are not significantly different from .42 and −.29. Thus, it seems the data cannot provide a definite answer to the issue involved—nor, probably, can any correlational study of this nature.

Third, if a timed aptitude test arouses stress cues, which elicit anxiety responses which affect the aptitude test performance, then a correlation between an anxiety measure and an aptitude measure could be interpreted to mean that the anxiety scale measures the anxiety which is a component in the aptitude test situation and which is reflected in the aptitude test scores. If this were true, then it would follow that an *aptitude* measure would correlate more highly with an anxiety-provoking college examination than with one that aroused little anxiety. Because the final examinations require more involvement on the part of the student and count more toward his total grade, these tests should be more anxiety-provoking than the midterm examination. From this reasoning, these two examinations were compared to provide the anxiety difference against which to test the aptitude measure. In no sample was there a significantly greater correlation between aptitude and final examination than between aptitude and midterm examination. These findings constitute a weak refutation of the third explanation.

A more definitive technique for evaluation of this explanation would be to administer the aptitude test to a group of students before and after they had been subjected to a course of psychotherapeutic treatment directed at the reduction of examination anxiety. If the increase in the mean aptitude score for the group receiving psychotherapy were significantly larger than any increase in the control group which had not received the psychotherapy, then we would have strong support for the third explanation of the relationship between aptitude and anxiety. This experiment would be similar to that by Pomeroy (1950), who demonstrated increases in finger-maze performance after only 10 min. of cathartic-type therapy. Until such time as appropriate data are available, however, one must hold in abeyance any final conclusions regarding the validity of this explanation.

In summary, regarding these three explanations, the first is refuted, doubt cast on the second, and the third neither refuted nor supported. Present evidence allows one to say only that either the third or the fourth explanation (that is, that the obtained correlation is a measure of the relationship between the two constructs—anxiety and aptitude) may be correct, or that both of them may be true. Had the first and second explanations been supported, they would support a preference for the use of a general anxiety scale because it is free of the contaminating influence of aptitude. The last two explanations support a preference for a specific anxiety scale because of its greater sensitivity. Since the evidence does indicate that these explanations are more probably correct, one has further support beyond that derived from the analysis of the interrelationships of the anxiety scales for the use of the specific rather than the general anxiety scales for prediction in academic performance situations.

DIRECTION OF EFFECT

A study of the anxiety literature leads to the question of how anxiety, when aroused in an examination, will affect performance; that is, whether it will facilitate it, debilitate it, or perhaps have no effect on it at all. On the basis of their earlier theoretical considerations, Sarason, Mandler, and Craighill (1952) state the following hypothesis:

> When a stimulus situation contains elements which specifically arouse test or achievement anxiety, this increase in anxiety drive will lead to poorer performance in individuals who have test-irrelevant [incompatible or interfering] anxiety responses in their response repertory. For individuals *without such response tendencies*, [italics added] these stimulus elements will raise their general drive level and result in improved performance (p. 561).

The implication is that measurable anxiety responses, when present, are debilitating to performance, an implication which is reflected

in the Mandler-Sarason TAS. All the items of the TAS are unidimensional, i.e., anxiety responses are either debilitating or not (e.g., "While taking a course examination, to what extent do you worry? Worry a lot. . . [to]. . . Worry not at all." [Item No. 35]). From the absence of negative responses for an *S*, that is, a low score on the test, they infer that when anxiety-provoking cues are present in the environment, "these stimulus elements will raise [his] general drive level and result in improved performance for that *S*." (Sarason, Mandler, & Craighill, 1952, p. 561). This results in their confounding the two alternatives of a facilitating effect of anxiety and no effect of anxiety on academic performance and leads to their failure to allow for a third alternative, the possible existence of an individual whose anxiety responses in an anxiety-provoking situation do not affect test performance either by improving or by depressing the score. Such an individual would have to be described in their formulation as one whose negative or debilitating response effect is just counteracted by the increased task-performance drive.

There is another approach which does not require this unwarranted inference and according to which an anxiety scale would measure the presence and intensity of both kinds of anxiety responses, those which facilitate performance and those which interfere with it. If an individual scored low on both of these scale components, he would be considered to be unaffected (insofar as test performance was concerned) by anxiety-provoking cues. In other words, the facilitating effects of anxiety would then be measured independently and not inferred from the absence of negative responses. This independent measure would allow for the possibility of the absence of both types of response as well as for the presence of either one.

Theoretically, Mandler and Sarason would predict, on the basis of their ideas, that two such independent measures of anxiety would be highly negatively correlated. That is, if an individual had a great deal of facilitating anxiety, he would have little debilitating anxiety and vice versa. This is the assumption underlying their scale because it allows them to use only the debilitating index and to infer facilitating anxiety from the absence of debilitation. This assumption may be unnecessary. In fact, these two constructs of debilitating and facilitating anxiety may be uncorrelated. Thus, an individual may possess a large amount of both anxieties, or of one but not the other, or of none of either. The nature of this correlation can be determined empirically following the construction of two such independent measures of anxiety, a scale measuring the facilitating effects of anxiety on achievement performance, and a separate scale measuring the debilitating effects of anxiety on achievement performance.

The Achievement Anxiety Test (AAT) was constructed to make this determination. It consists of two independent scales: a facilitating scale of nine items based on a prototype of the item—"Anxiety helps me to do better during examinations and tests"; and a debilitating scale of 10 items based on a prototype of the item—"Anxiety interferes with my performance during examinations and tests." Both scales have gone through numerous revisions based upon item analyses, correlations with various criteria, and theoretical reformulations. The test-retest reliabilities for a 10-week interval are .83 and .87, respectively. The test-retest reliability over an 8-month period is .75 for the facilitating scale and .76 for the debilitating scale. The two scales are administered in one questionnaire, the items randomly mixed (as presented below). The *S*s answer each item on a five-point scale, indicating the degree to which the item applies to them. The AAT includes neutral buffer items in addition to the items listed. The numbers in the parentheses to the right are the actual item numbers on the AAT.

Facilitating Anxiety Scale

1. I work most effectively under pressure, as when the task is very important. Always—Never. (2)
2. While I may (or may not) be nervous before taking an exam, once I start, I seem to forget to be nervous. I always forget—I am always nervous during an exam. (9)
3. Nervousness while taking a test helps me do better. It never helps—It often helps. (11)
4. When I start a test, nothing is able to distract me. This is always true of me—This is not true of me. (12)
5. In courses in which the total grade is based mainly on *one* exam, I seem to do better than other people. Never—Almost always. (14)
6. I look forward to exams. Never—Always. (16)
7. Although "cramming" under pre-examination tension is not effective for most people, I find that if the need arises, I can learn material immediately before

an exam, even under considerable pressure, and successfully retain it to use on the exam. I am always able to use the "crammed" material successfully—I am never able to use the "crammed" material successfully. (19)

8. I enjoy taking a difficult exam more than an easy one. Always—Never. (21)

9. The more important the exam or test, the better I seem to do. This is true of me—This is not true of me. (24)

Debilitating Anxiety Scale

1. Nervousness while taking an exam or test hinders me from doing well. Always—Never. (1)

2. In a course where I have been doing poorly, my fear of a bad grade cuts down my efficiency. Never—Always. (3)

3. When I am poorly prepared for an exam or test, I get upset, and do less well than even my restricted knowledge should allow. This never happens to me—This practically always happens to me. (5)

4. The more important the examination, the less well I seem to do. Always—Never. (6)

5. During exams or tests, I block on questions to which I know the answers, even though I might remember them as soon as the exam is over. This always happens to me—I never block on questions to which I know the answers. (10)

6. I find that my mind goes blank at the beginning of an exam, and it takes me a few minutes before I can function. I almost always blank out at first—I never blank out at first. (15)

7. I am so tired from worrying about an exam, that I find I almost don't care how well I do by the time I start the test. I never feel this way—I almost always feel this way. (17)

8. Time pressure on an exam causes me to do worse than the rest of the group under similar conditions. Time pressure always seems to make me do worse on an exam than others—Time pressure never seems to make me do worse on an exam than others. (18)

9. I find myself reading exam questions without understanding them, and I must go back over them so that they will make sense. Never—Almost always. (23)

10. When I don't do well on a difficult item at the beginning of an exam, it tends to upset me so that I block on even easy questions later on. This never happens to me—This almost always happens to me. (26)

Part of the refinement of the scales was accomplished through means of empirical techniques. After a large number of items had been constructed for each scale, the scales were used to predict several performance criteria, such as grade-point averages and final examination grades. These data were item-analyzed to give the correlation of each item with the criterion. Those items were retained which were highly correlated with the criteria but which were not correlated with each other. In this way, it was hoped to minimize the intercorrelation of the scales without affecting their validity coefficients. The final correlations between the facilitating and debilitating scales were −.37, −.34, −.43, and −.48, drawn from four different samples. For these four samples, the average means and standard deviations for the facilitating and debilitating scales of the AAT were as follows: for the AAT+, $M = 27.28$, $SD = 4.27$; for the AAT−, $M = 26.33$, $SD = 5.33$. The average correlation for the combined N (379) was −.37. All these correlations are significant beyond the 1% level. Hence, in spite of efforts to separate the two scales empirically, a low but significant correlation remains. The point seems made, however, that the correlation is neither perfect nor high as Mandler and Sarason imply and require as an assumption for the use of their scale.

In general, the two scales of the AAT correlated about equally, though opposite in sign, with the other variables discussed in this paper (see Tables 1, 2, and 3). The important question is whether, when predicting an academic performance score such as college grade-point average, there is any advantage in using both scales rather than using merely a conventional debilitating anxiety scale such as the AAT− alone or the Mandler and Sarason TAS. These two measures of debilitating anxiety (the AAT− and the TAS) correlate .64 with an N of 40, indicating considerable equivalence. Table 6 answers the question of whether a scale which measures the facilitating effect of anxiety justifies itself. In the last two columns to Table 6 it can be seen that for three separate samples, the multiple correlations, using both the plus and minus scales to predict grade-point average, are signifi-

TABLE 6

MULTIPLE CORRELATIONS PREDICTING GRADE-POINT AVERAGE (GPA) FROM THE AAT

N	R^a	r^b	r^c	r^d	p^e	p^f
93	.50	.36	−.45	−.37	.001	.05
92	.32	.32	−.08	−.34	ns	.01
96	.54	.50	−.40	−.43	.05	.001

[a] GPA · AAT+ & AAT−.
[b] GPA · AAT+.
[c] GPA · AAT−.
[d] AAT− · AAT+.
[e] Significance of adding AAT−
[f] Significance of adding AAT+.

cantly better predictors than the minus or the plus scales alone.

It seems clear, therefore, that the incorporation of items designed to measure facilitating anxiety into a scale which already effectively measures debilitating anxiety can significantly increase the prediction of academic performance scores.

SUMMARY

The results support the following conclusions:

1. Specific anxiety scales (scales having items specific to the academic test situation) and general anxiety scales are measuring, to a significant extent, something different from one another.

2. The specific anxiety scales are better predictors of academic performance than are the general anxiety scales.

3. The general anxiety scales are not significantly related to verbal aptitude, while the specific anxiety scales are all related to aptitude.

4. The specific anxiety scales, although more highly correlated with aptitude than the general anxiety scales, are, nevertheless, more often than the general anxiety scales, able to account for variance in academic performance other than that accounted for by a measure of aptitude.

5. An explanation of the relationship between anxiety scales and measures of aptitude, based on the use of an anxiety scale as a vehicle for rationalization or justification by students with poor past performance in academic situations, is not supported by these data.

6. On the basis of these data, no conclusion may be reached about explanations of the relationship between anxiety scales and measures of aptitude on the assumption that aptitude scores are, themselves, affected by anxiety. A possible experiment which may clarify this issue is suggested.

A scale for the measurement of "facilitating" anxiety was constructed, which added significantly to the prediction of grade-point average when it was combined with a measure of "debilitating" anxiety. The theoretical and practical significance of these findings was discussed.

REFERENCES

ALPERT, R. Anxiety in academic achievement situations: Its measurement and relation to aptitude. Unpublished doctoral dissertation, Stanford Univer., 1957.

CHILD, I. L. Personality. *Annu. Rev. Psychol.*, 1954, **5**, 149–170.

COLLEGE ENTRANCE EXAMINATION BOARD. *A description of the College Board Scholastic Aptitude Test.* Princeton: Educ. Testing Serv., 1956.

DAVIDS, A. Relations among several objective measures of anxiety under different conditions of motivation. *J. consult. Psychol.*, 1955, **19**, 275–279.

DAVIDS, A., & ERIKSEN, C. W. The relation of manifest anxiety to association productivity and intellectual attainment. *J. consult. Psychol.*, 1955, **19**, 219–222.

FREEMAN, M. J. The development of a test for the measurement of anxiety: A study of its reliability and validity. *Psychol. Monogr.*, 1953, **67**(3, Whole No. 353).

MANDLER, G. Anxiety and performance: Empirical and theoretical aspects of anxiety scales. Unpublished manuscript, Lab. Soc. Relat., Harvard Univer., 1954.

MANDLER, G., & SARASON, S. B. A study of anxiety and learning. *J. abnorm. soc. Psychol.*, 1952, **47**, 166–173.

PECK, RUTH. The influence of anxiety factors upon the effectiveness of an experimental "counselling session." Unpublished doctoral dissertation, State Univer. Iowa, 1951.

POMEROY, D. S. Ameliorative effects of "counselling" upon maze performance following experimentally induced stress. *Amer. Psychologist*, 1950, **5**, 327. (Abstract)

SARASON, S. B., & MANDLER, G. Some correlates of test anxiety. *J. abnorm. soc. Psychol.*, 1952, **47**, 810–817.

SARASON, S. B., MANDLER, G., & CRAIGHILL, P. G. The effect of differential instructions on anxiety and learning. *J. abnorm. soc. Psychol.*, 1952, **47**, 561–565.

SCHULZ, R. E., & CALVIN, A. D. A failure to replicate the finding of a negative correlation between manifest anxiety and ACE scores. *J. consult. Psychol.*, 1955, **19**, 223–224.

SPENCE, K. W., & TAYLOR, JANET A. Anxiety and strength of the USC as determiners of the amount of eyelid conditioning. *J. exp. Psychol.*, 1951, **42**, 183–188.

TAYLOR, JANET A. The relationship of anxiety to the conditioned eyelid response. *J. exp. Psychol.*, 1951, **41**, 81–92.

TAYLOR, JANET A. A personality scale of manifest anxiety. *J. abnorm. soc. Psychol.*, 1953, **48**, 285–290.

TAYLOR, JANET A., & SPENCE, K. W. The relationship of anxiety level to performance in serial learning. *J. exp. Psychol.*, 1952, **44**, 61–64.

WELSH, G. S. An anxiety index and an internalization ratio for the MMPI. *J. consult. Psychol.*, 1952, **16**, 65–72.

(Received May 21, 1959)

Journal of Abnormal and Social Psychology
1960, Vol. 61, No. 2, 216-222

EFFECTS OF WAITING WITH OTHERS ON CHANGES IN LEVEL OF FELT ANXIETY[1]

LAWRENCE S. WRIGHTSMAN, Jr.[2]
University of Minnesota

AN OLD familiar saying has it that "misery loves company." This having been verified empirically by Schachter (1959), a next question might simply be "Why?" Why is it that most people, ordinarily content to be left alone, become seekers of others when they become upset, moody, or threatened? Why does an increase in a person's level of anxiety cause him to be more likely to want to be with others?

Several possible reasons for the relationship between anxiety and the affiliative tendency come easily to mind, including the expectation of irrelevant and diverting conversation with others and the idea of "safety in numbers." But a series of analytic experiments by Schachter rule out some of these as valid explanations. Schachter concludes that an anxious subject's desire to be with others is determined by (*a*) needs for direct reduction of anxiety, and/or (*b*) needs for evaluation of one's emotional state. The present study examines some of the implications of the relationship between anxiety and affiliation.

The first explanation implies that people want to be with others when anxious because they expect from others sympathy and reassurance that is directly anxiety-reducing. ("Anxiety" refers here to a feeling of uneasiness or concern about participating in an impending unpleasant event; unlike generalized anxiety or manifest anxiety it is conceived as linked to a specific situation.) According to the first hypothesis of the study, then, if persons are anxious about an impending event, being with others is more effective in reducing their anxiety than is being alone.

Despite the widely held belief in the effectiveness of the presence of other people in moderating one's extreme fears, there is no direct evidence that other people's statements facilitate the reduction of anxiety about a specific situation. Indirect evidence, however, comes from Schachter's study (1959). His subjects were given a description of the "electric shock experiment" in which they were to participate. By varying description of the anxiety-producing features of the "experiment" and the alternatives involved in a short waiting period before anticipated participation in it, Schachter found that *S*s who expected a very painful experiment more frequently chose to wait with others than did their counterparts who were led to believe that the experiment would be painless. Further evidence is afforded by the study of Davitz & Mason (1955), who found that the presence of one rat tended to reduce the strength of a fear response in another rat.

The second hypothesized explanation for the link between anxiety and affiliation implies that the presence of others gives a person a chance to evaluate his own level of emotion by comparing it with that of others. Perhaps the situation is so novel or ambiguous that a person is unclear about the appropriate emotional response. In such a case the responses of others may serve to indicate the proper level of concern.

This interpretation would seem to follow from an extension to the evaluation of emotions of the scheme developed by Festinger (1954) for explaining how certain processes of social comparison operate in the evaluation of an ability or an opinion. (The theory has been verified empirically by Dreyer, 1954; Gerard, 1953.) Festinger hypothesizes that there is a drive to evaluate one's own abilities and opinions, and when no objective method of

[1] This study is a condensation of a dissertation submitted in partial fulfillment of the requirements for the PhD degree at the University of Minnesota. The writer wishes to express his appreciation to Stanley Schachter for his guidance in the study. This experiment is part of a program of research on Social Determinants of Emotional State which is being carried out at the Laboratory for Research in Social Relations under the direction of Schachter and the sponsorship of the National Institute for Mental Health of the United States Public Health Service.

[2] Now at George Peabody College for Teachers.

evaluation is possible, the person uses social reality—the statements and behaviors of others—to aid in his self evaluation. For social comparison to be most informative, the person's comparison objects should be other persons whose levels of ability or opinion are similar to his own. The process of social comparison thus gives rise to pressures toward uniformity, by which the levels of others are brought closer to the person's own level.

According to Festinger, these pressures toward uniformity result in three maneuvers: (*a*) change in the person's own ability or opinion so as to bring it closer to that of the comparison persons, (*b*) the attempt to move others toward one's own level, and (*c*) cessation of comparison of one's own position with those of persons whose abilities or opinions differ widely from one's own.

One of the purposes of this study is to determine whether the emotion of anxiety is similar to abilities and opinions in being evaluated in ways subject to influence by the responses of others. If the same social comparison processes obtain, the same maneuvers should be found in a group of people placed in a sudden anxiety-producing situation that emerge in people seeking to evaluate their opinions. The second hypothesis of this study is, then, that pressures toward uniformity in level of anxiety, namely influence processes and rejection tendencies, develop when groups of persons are placed in an ambiguous emotional situation.

But is it correct to assume that felt emotions are, in part, cognitively determined, and that there is a conceptual similarity between the determinants of emotions and the determinants of opinions? The tenor of much thinking about the emotions seems to suggest that both physiological and cognitive processes determine experienced emotions. Ruckmick (1936), for example, states that "there is a tendency . . . to admit that purely bodily expression without the cognitive or 'intellectual' factor does not qualify as an equation for the experienced real emotion," (p. 343) and that "What the child . . . feels depends more on what he cognizes or, in short, on the accumulations of his perceptual and ideational experiences," (p. 448). Young (1943) has a similar point of view. To the extent that cognitive factors do determine emotional states, emotions, like opinions, may be expected to be susceptible to social influence.

METHOD

Subjects

The subjects (*Ss*) were 204 students, 120 females and 84 males, enrolled in undergraduate psychology classes. By participating the students earned bonus points applied to their examination scores.

Experimental Session

Four *Ss* of the same sex were used in each experimental session. When each *S* arrived, *E* or his nurse assistant[1] placed him in a separate room in which *S* found a tray containing hypodermic needles, syringes, cotton, alcohol, and other injection equipment. After completing a few biographical questions, including several dealing with his physical health, each *S* heard over the intercommunication system a tape recording describing the experiment in which he was to participate. The experiment was described as one dealing with glucose level and mental activity. *S* was told that his glucose level was to be drastically changed by a series of injections of either a depressant drug or a glucose additive. Because these substances had to be injected directly into the bloodstream, it was likely that they would be more painful than routine injections. *S* was then warned that the injection of the depressant drug might make him feel drowsy and lacking in energy but that he should be able to respond. If he was given the additive, it would raise his glucose level to an absolute maximum amount near the point which brings on glucose-shock, fainting, and similar symptoms of hyperglycemia or diabetes. A quick blood test to determine glucose level would be given by pricking a finger. All the above operations were to occur in private.

Immediately after hearing this, each *S* was asked to indicate on a formsheet how at-ease or ill-at-ease he felt about participating in the described experiment by writing a number from 0 (completely at-ease) to 100 (completely ill-at-ease).

At this point the procedure was varied for the three conditions of the experiment. Two Together conditions were used, along with an Alone condition. Verbal communication was forbidden in one Together condition, but allowed to run its course in the other, to permit measuring the possible effects of verbal communication on reduction of anxiety. All four *Ss* in a particular experimental session were in the same condition, the particular condition being determined at random.

If the *Ss* were in the Together-Talk (T-T) condition, they were brought to a large waiting room and told:

You are supposed to wait here until we are ready

[1] The author wishes to thank Bernice P. Flynn, who served authentically as the nurse assistant.

TABLE 1
MEAN RESPONSES AND CHANGES IN MEASURES OF ANXIETY
(Data for 17 groups in each condition, 68 *S*s in all)

Condition	Prewaiting-Period Measure	Postwaiting-Period Measure	Change from Pre- to Postwaiting-Period Measure
Together-Talk	43.7	37.3	−6.4
Together-No-Talk	42.8	36.6	−6.2
Alone	41.3	37.2	−4.1

to use you. You can read or smoke or study if you like. If you want to talk, that's okay. You can talk about the experiment or anything you like. There will be only a short wait.

If the *S*s were to be in the Together-No-Talk (T-N-T) condition, they were brought to the same room and told:

You are supposed to wait here until we are ready to use you. You can read or smoke or study if you like. However, please do *not* talk. Part of the experiment is of a verbal nature and we want only your own ideas to be given. Therefore, it is important that you not talk about the experiment or anything else. There will be only a short wait.

If the four *S*s were in the Alone condition, each was left in his separate room and told:

You are supposed to wait here until we are ready to use you. You may read or smoke or study if you like. There will be only a short wait.

Immediately after this announcement, *E* and his assistant left the room for 5 min. *S*s in the Together conditions were observed through a one-way screen. All *S*s in the T-N-T condition followed the restriction against talking. After 5 min. the assistant returned and told the *S*(s), "We are about ready for you but before we begin I need to get some more information from you." She then gave each *S* a set of questions which consisted of a repetition of the prewaiting-period measure and other questions dealing with *S*'s concern about the experiment.

When they completed these, *S*s were brought together if they had been separated and a "catharsis" session was held, in which the true purpose of the experiment and the need for deception were explained. *S*s' questions were answered, and they agreed to keep the true nature of the experiment secret. Each *S* then completed a questionnaire dealing with his experimental background, his suspicions about the described experiment, and the extent of his friendship with the other *S*s at that session.

RESULTS

Effect of Being with Others on Reduction of Anxiety

If the first hypothesis is correct, *S*s waiting together should show greater reduction of anxiety than do *S*s waiting alone. Table 1 presents the mean responses for the pre- and postwaiting-period measures for *S*s in the three conditions. Means for the three conditions are quite similar on all measures. None of the differences between conditions is statistically significant, nor are differences between Alone condition *S*s and all Together condition *S*s pooled together. In each condition the average change was in a downward direction, from four to six points. When all conditions are combined, the mean level of anxiety of 37.0 on the postwaiting-period measure was significantly less ($p < .001$) than the mean of 42.5 on the same measure before the waiting period.

The number of *S*s wishing to withdraw from the experiment without receiving experimental credit was also similar in each condition. That 26 of the 203 *S*s responding to this question wished to withdraw, gives some indication of the effectiveness of the described experiment in creating anxiety.

The nonsignificant differences pertaining to the first hypothesis give no support for the belief that being with others directly reduces anxiety. But Schachter (1959) found that the relationship between anxiety and the desire to be with others was almost entirely restricted to those *S*s who were first or only children. Accordingly, such *S*s might be affected more strongly than others by the social situation. Table 2 presents the results of an analysis by birth order.

First-born *S*s in each of the Together conditions showed a greater net change than did first-born *S*s in the Alone condition. In both the Together-vs.-Alone comparisons, differences are significant at the .05 level, using a two-tailed Mann-Whitney *U* test because of the heterogeneity of variances. Thus, it appears that for *Ss who are first children or only children*, being with others is indeed effective in reducing anxiety over an impending unpleasant event.

In the case of the *S*s in Table 2 who were later children, however, the mean net change does not differ from one condition to another. No mean differences are statistically significant. For these *S*s being with others appears not to have reduced anxiety.

Table 2 introduces a new measure, that of

TABLE 2
MEAN RESPONSES ON MEASURES OF ANXIETY ACCORDING TO BIRTH ORDER OF *S*

Condition (First-born and only-child *Ss*)	*N*	Prewaiting-Period Measure	Postwaiting-Period Measure	Net change, Pre- to Postwaiting Period	Gross Change, Pre- to Postwaiting Period
T-T[a]	32	39.8	31.8	−8.0[b]	10.8[c]
T-N-T[d]	33	43.2	33.9	−9.3[b]	9.9[b]
Alone[a]	41	42.1	38.6	−3.5	4.2
(Later-children *Ss*)					
T-T[a]	35	47.8	42.9	−4.9	6.6
T-N-T[d]	35	42.4	39.1	−3.3	4.2
Alone[a]	26	39.9	35.4	−4.5	6.0

[a] Birth order of one *S* in T-T (Together-Talk) condition and one *S* in Alone condition were not recorded.
[b] Significantly different from comparable Alone condition mean at .05 level, using Mann-Whitney *U* test.
[c] Significantly different from comparable Alone condition mean at .01 level, using Mann-Whitney *U* test.
[d] Together-No-Talk.

"gross change," or the absolute magnitude of change disregarding direction. For this measure the same tendency is found. First-born *Ss* waiting together show significantly greater gross change than first-born *Ss* waiting alone; again in the case of later-children *Ss* there are no differences. There is a notable difference between the first children in the T-T and T-N-T conditions: those in the T-T condition show greater gross change but less net change than do first-children *Ss* in the T-N-T condition. A comparison of first borns and later children *within* the T-T condition indicates that they differ significantly in gross change but not in net change. On the other hand, in the T-N-T condition first-born *Ss* show significantly greater change, both gross and net, than do later-children *Ss*. It thus appears that verbal communication may actually lead to an increase in anxiety.

On the remaining postwaiting-period measures of anxiety, the between-condition differences in birth order were similar to those already presented.

Effect of Being with Others on Self-Evaluation

If others are used as objects of social comparison in a group situation, changes in a person's self-ascribed level of anxiety and influence on that of others should occur. According to the second hypothesis of this study, then, interindividual variability in level of anxiety should be reduced during the waiting period for groups of four people waiting *together*. Such homogenization should not occur in sets of four *Ss* each waiting alone. The four *Ss* at each session were treated as a unit, and the ratio of the range after the waiting period divided by the range before the waiting period was used as a measure of homogenization. The second hypothesis then holds that lower average range ratios should occur in the Together conditions than in the Alone condition.

The results confirm the expectation. The mean homogenization ratios in both of the Together conditions are significantly less ($p < .05$) than the mean ratio for the Alone condition (T-T, 17 sets, .83; T-N-T, .84; Alone, .98). The Alone condition shows a level of homogenization that does not differ significantly from the expected ratio of 1.0, while in the Together conditions the ratios are significantly smaller than 1.0 at the .02 level or better.

Since the mean initial ranges for the three conditions were quite similar, the differences between conditions cannot be attributed to regression effects. But in the two Together conditions other possible artifactual causes of homogenization should be considered, particularly those related to the general tendency toward reduction of anxiety and to "floor" effects associated with the zero point of the scale. This kind of artifact should have the

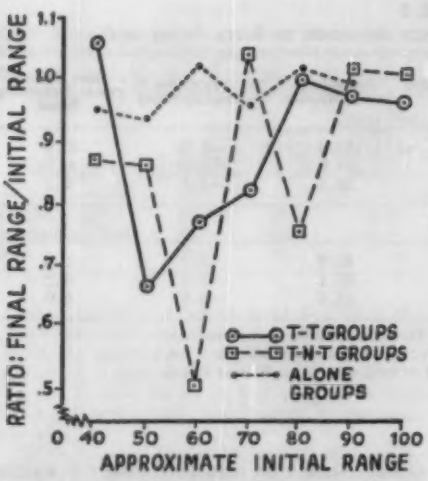

FIG. 1. The effect of initial range on extent of homogenization. (T-T refers to the Together-Talk condition; T-N-T to Together-No-Talk.)

effect that groups in which anxiety was substantially reduced should also show a high degree of homogenization; there should be a sizable correlation between the amount of reduction in anxiety and the measure of homogenization. In the T-T condition this correlation is +.31 and nonsignificant, indicating that any such artifact has a minimal effect. But in the T-N-T condition it is + .57, which is significant at the .02 level. Part of the homogenization of levels of anxiety in the T-N-T condition thus appears to be attributable to the general reduction in anxiety and the floor set by the zero point of the scale.

To the extent that interindividual variability in anxiety is nonartifactually reduced, indirect evidence is provided for the first two of Festinger's hypothesized maneuvers to reduce discrepancies—change in the person's own responses and the attempt to change the responses of others. If *S*s employ processes of social comparison to facilitate self-evaluation of their emotional states, evidence should be present in the T-T condition for Festinger's third hypothesized maneuver—cessation of comparison with those who differ widely from the person's own position. In the T-T condition, this effect should emerge more often in groups in which the initial range in levels of anxiety is great. In such groups, from the perspective of a given member, other group members should often appear so deviant and so resistant to change that they are not within the limits necessary for social comparison. There should thus be less homogenization of levels of anxiety in such groups than in the groups with moderate initial range.

Few pressures toward uniformity should be expected in T-T groups with a narrow initial range because the members are already quite uniform in level of anxiety; less homogenization of levels of anxiety may therefore be expected in such groups than in those with a moderate initial range. In groups under the Alone condition, on the other hand, homogenization should be unrelated to initial range since there is no opportunity for social comparison to occur. The T-N-T condition is more difficult to predict but should resemble the T-T-condition results to some degree if processes of social comparison are operating.

Figure 1 indicates that the data are in agreement with these predictions. (Because the initial range for most groups was a multiple of five, ranges have been grouped together in Fig. 1. From one to five groups are represented by each point.) The three T-T groups with initial ranges of 90 points or greater and the two groups with an initial range of 40 points or less showed less homogenization than did the 12 groups with moderate initial ranges. (Of the latter groups, five had a range of approximately 50 points, one 60 points, five about 70 points, and one of 80 points.) No such pattern is evident for the sets in the Alone condition, where homogenization is unrelated to initial range. The results for the T-N-T condition were more erratic, but their resemblance to those for the T-T condition tentatively suggests that comparison processes may be active in the absence of verbal communication.

A possible artifactual cause for the differences in homogenization between wide-range and moderate-range groups might lie in a tendency for those *S*s whose initial reaction was extreme (either less than 25 or greater than 75) to be more resistive to change, but a

further analysis of the data gives no support to this possibility.

DISCUSSION

Effect of Being with Others on Reduction of Anxiety

The differences according to birth order in the anxiety-reducing effect of being with others fit in neatly with Schachter's finding (1959) that first-born *Ss* have a greater desire to be with others when anxious. It now appears that their reason for this greater desire can be attributed to the fact that being with others directly reduces their anxiety. Later-children *Ss*, for whom being with others does not have this effect, therefore have no particular desire to be with others when anxious.

The results for the T-N-T condition are consistent with another of Schachter's findings, that first-born *Ss* desire to be with others when anxious even when verbal communication is not permitted. The present study finds, correspondingly, that when such *Ss* *are* placed in a situation without verbal communication, where only the sheer physical presence of others is available, they respond with a greater reduction in anxiety than occurs in similar *Ss* who remain alone.

But why this birth-order effect? The cause is certainly not a contemporary one; it must be traced to childhood experiences. There are indications in the literature that first and later children differ in various ways in their early environment. Studies of the parental treatment of different siblings (Gesell & Ilg, 1942; Sears, Maccoby, & Levin, 1957) indicate that the first child receives more attention and nurturance from his parents than do his later siblings. Sibling rivalry, with the displacement of hostilty upon the younger child (Baldwin, 1955), appears to occur frequently. On the basis of present knowledge, however, there is no compelling reason for preferring any one of several possible explanations, which invite further research.

Effect of Being with Others on Self-Evaluation

The strong support given to the second hypothesis indicates that Festinger's scheme for explaining self-evaluation through a social comparison process is useful for the emotion of anxiety as well as for opinions and abilities. Can it then be concluded that because cognitive factors have some effect in determining this particular emotional state that the emotions in general are also susceptible to social influence? Some evidence for an affirmative answer is found in a study which demonstrates that another emotion—the state of boredom or satiation—can be influenced by processes of social comparison (Horwitz, Exline, Goldman, & Lee, 1953). In this investigation of the effects of different group characteristics on task satiation, it was found that in the one condition where verbal communication was pres[illegible] the variability among the members' levels of satiation was much less than the variability within the no-communication conditions.

SUMMARY

Two determinants for the relationship between anxiety and the affiliative tendency have been proposed. These are the need for direct reduction of anxiety and the need for self-evaluation of level of anxiety. The purpose of this study was to examine some of the implications of these two explanations. It was hypothesized that (*a*) being with others would be more effective in reducing anxiety than would being alone, and (*b*) pressures toward uniformity in level of anxiety would be elicited in groups of persons placed in an ambiguous emotional situation, as a consequence of requirements for self-evaluation of level of anxiety.

The first hypothesis was not confirmed, but an alternative hypothesis—that for first-born *Ss*, being with others is more effective in reducing anxiety than is being alone—received strong support.

The second hypothesis received strong support. Pressures toward uniformity in regard to level of anxiety emerged in *Ss* waiting together. It was concluded that a person does evaluate his level of anxiety through the process of social comparison and that level of anxiety shares with opinion the feature of being partially determined by cognitive processes.

REFERENCES

BALDWIN, A. L. *Behavior and development in childhood.* New York: Dryden, 1955.

DAVITZ, J. R., & MASON, D. J. Socially facilitated reduction of a fear response in rats. *J. comp. physiol. Psychol.*, 1955, **48**, 149–151.

DREYER, A. S. Aspiration behavior as influenced by expectation and group comparison. *Hum. Relat.*, 1954, **7**, 175–190.

FESTINGER, L. A theory of social comparison processes. *Hum. Relat.*, 1954, **7**, 117–140.

GERARD, H. B. The effects of different dimensions of disagreement in the communication process in small groups. *Hum. Relat.*, 1953, **6**, 249–272.

GESELL, A., & ILG, FRANCES L. *Infant and child in the culture today: The guidance and development in home and nursery school.* New York: Harper, 1942.

HORWITZ, M., EXLINE, R. V., GOLDMAN, M., & LEE, F. J. Motivational effects of alternative decision-making processes in groups. *ONR tech. Rep.*, 1953. (Contract N6ori-07144, Group Psychol. Branch)

RUCKMICK, C. A. *The psychology of feeling and emotion.* New York: McGraw-Hill, 1936.

SCHACHTER, S. *The psychology of affiliation.* Stanford, Calif.: Stanford Univer. Press, 1959.

SEARS, R. R., MACCOBY, ELEANOR, & LEVIN, H. *Patterns of child rearing.* Evanston, Ill.: Row, Peterson, 1957.

YOUNG, P. T. *Emotion in man and animal.* New York: Wiley, 1943.

(Received June 24, 1959)

Journal of Abnormal and Social Psychology
1960, Vol. 61, No. 2, 223-230

THE INFLUENCE OF SOME TYPES OF POWER RELATIONSHIPS AND GAME STRATEGIES UPON THE DEVELOPMENT OF INTERPERSONAL TRUST[1]

LEONARD SOLOMON[2]

Research Center for Human Relations, New York University

THIS study is part of a research program developed by Deutsch (1957), which sought to investigate the personal and situational determinants of interpersonal trust. In his initial studies Deutsch examined the influence of cooperative, individualistic, and competitive motivational orientations upon the development of mutual trust in a two-person bargaining game. A bargaining relationship is one in which the participants have a conflict of interests yet are dependent upon each other in order to avoid a mutually detrimental outcome (Schelling, 1958). One rather striking finding to emerge from Deutsch's study was the fact that under conditions of noncommunication the individualistically oriented dyads did not differ from the competitive ones with respect to the number of trusting choices made in a ten-trial game. Contrary to what would have been predicted on a common sense basis, the individualistic subjects (*S*s) were unable to act rationally in the absence of mutual trust. Each player was unable to subordinate immediate self-interest for long range group gain and the relationship tended to deteriorate into a self-defeating competition. These findings stimulated several follow-up studies that attempted to determine the minimal social or situational conditions under which mutual trust could arise in dyads with an individualistic orientation. Thus, Loomis (1959) examined the effects of variations in the opportunity to communicate while Farr (1957) dealt with the influence of a mutually disliked third person upon the development of mutual trust.

The present study sought to determine how a player in different positions of relative power could exercise this power in the form of various game playing strategies so as to induce an individualistic *S* to adopt a cooperative orientation. Deutsch's formulations concerning the situational conditions necessary for the realization of mutual trust, along with Heider's principle of cognitive balance (Heider, 1958) provided a theoretical basis for the derivation of several hypotheses.

Trust and suspicion. For the purposes of economy and clarity of exposition, the symbol *S* is used to refer to the perceiver (also the subject) and his experiences or cognitions; *O* refers to another person whose action affects *S*. In the terms of the dyadic game employed in this study, *S* refers to Player 1 and *O* refers to Player 2.

Deutsch (1958) has proposed the following conceptualizations of interpersonal trust and suspicion:

A person (*S*) may be said to have trust in the occurrence of an event, if *S* expects its occurrence and his expectation leads to behavior that he perceives to have positive motivational consequences if the expectation is confirmed and negative motivational consequences if it is not. *S* may be said to be suspicious of the occurrence of an event if the disconfirmation of the expectation of the event's occurrence is preferred to its confirmation, and if the expectation of its occurrence leads to behavior that is intended to reduce its negative motivational consequences.

Interpersonal trust or suspicion exists when *S* trusts or suspects that *O* will do something and he perceives that *O* is aware that his behavior has relevance for *S*. Trust refers to expectations of benevolence; whereas

[1] This paper is adapted from a doctoral dissertation completed in 1957 at New York University. The author is deeply indebted to Morton Deutsch, under whose guidance this study was performed. The author also wishes to thank Robert S. Lee of New York University and Herbert E. Spohn, Harold Wilensky, and Jacob Cohen of the Montrose VAH research staff for their valuable suggestions with respect to the manuscript. This experiment was carried out as part of a research program on Conditions Affecting Cooperation and was supported by a grant from the Office of Naval Research under Contract NONR-285 (10). The United States government is authorized to reprint this article in part or whole.

[2] Now at Boston University.

suspicion refers to expectations of malevolence. Deutsch (1958) exemplifies these notions when he states (paraphrased):

If a mother trusts a baby sitter enough to leave her baby in her care and the sitter does not live up to this trust, the mother suffers an unpleasant consequence—harm to her baby. The trusting mother, when her trust is fulfilled, is in a more advantageous position than if she had not trusted. However, the disadvantages of trusting and having her trust unfulfilled are considerably greater than the advantages of being able to do things she could otherwise not do.

Cognitive balance. Heider (1958) has proposed that the process of social perception is steered by a need to achieve cognitively balanced states. The need for cognitive balance has been defined as a tendency for the person to cognize various aspects of his environment in such a way that the behavioral implications of his cognitions are not in contradiction. If an imbalanced state exists, the person will attempt to reduce or remove the imbalance through either cognitive reorganization or behavior. This principle and several of its derivations have been elaborated and tested by Festinger (1957) and Deutsch and Solomon (1959).

Interpersonal power. Heider has defined social power as the ability to produce intentionally events that have a particular degree of positive or negative motivational consequences for a recipient. In line with this conception the interpersonal power balance has been defined in this paper in the following ratio form: *O*'s ability to produce motivational consequences for *S* relative to *S*'s ability to produce motivational consequences for *O*. The nature of the events to be produced will determine what dimensions of ability required (i.e., physical strength or intelligence).

Game strategies. The two person bargaining games are so contrived that they represent a mixture of mutual and divergent interests. In such relationships the term *game strategy* is used to refer to the pattern of *O*'s choice behavior as it affects the preferences of *O* and *S*. As a first approximation, situations which result in both lesser loss as well as greater gain are treated, hereafter, as involving gain; loss is extended to include lesser gain as well as greater loss. The two modal forms of game strategy dealt with in this experiment are the Cooperative Strategy wherein *O*'s choice pattern promotes the maximum mutual gain, and the Noncooperative Strategy wherein *O* blocks the attainment of maximal mutual gain. The motivational source of such behavior patterns may vary. For example, *O*'s cooperation could stem from fear of retaliation (individualism) or a sincere liking for *O* (altruism). A guiding conception of this study is that the nature of the power relationship influences the kind of motive *S* attributes to account for *O*'s game strategy.

In terms of the game, *O*'s power over *S* is a reflection of his ability to achieve a preferred outcome relatively independent of *S*'s cooperation. Thus, if *O* had absolute power, he could employ a Noncooperative Strategy to obtain his optimal gain whereas if the players had equal power, *O* would have to reckon with potential losses derivable from *S*'s use of his retaliatory power. This reasoning suggests a more general proposition; the greater *S*'s power relative to *O* the more likely is *S* to expect cooperative as opposed to noncooperative behavior and to behave with trust rather than suspicion. In addition, the principle of cognitive balance serves to predict that an *S* who receives a Cooperative Strategy should tend to develop a liking for *O*, whereas, the contrary would be true if *S* received a Noncooperative Strategy. The intensity of the like or dislike is expected to be influenced by the particular power relationship that exists between *S* and *O*, or to put it another way, the perceived cost of cooperation or noncooperation to *O* influences *S*'s motive attribution and subsequent feelings toward *O*.

The specific hypotheses associated with this rationale are elaborated in the presentation of the results.

METHOD

A laboratory experiment was designed employing a two-person nonzero-sum game (see Fig. 1) in which the gains or losses incurred by each player were a joint function of the choices made between two alternatives by both players. Two sets of independent variables, namely, four types of power relationships and three types of game strategies, were created experimentally and analyzed as a 4 × 3 factorial design. A total of 256 undergraduate Queens College students served as *S*s. Thirty-six experimental groups each composed of

O in Absolute Power 1 (AP_1) (Coordinated Preferences)

	A	B
X	(+30, +30)	(−30, +30)
Y	(+20, +30)	(−10, +30)

O in Partial Power (PP)

	A	B
X	(+30, +30)	(−30, +40)
Y	(+20, +10)	(−10, +20)

O in Absolute Power 2 (AP_2) (Divergent Preferences)

	A	B
X	(+30, +30)	(−30, +40)
Y	(+20, +30)	(−10, +40)

O in Equal Power (EP)

	A	B
X	(+30, +30)	(−30, +40)
Y	(+40, −30)	(−20, −20)

(*S*) chooses between Rows X and Y, whereas (*O*) chooses between columns A and B. *S*'s payoffs are the first numbers in the parentheses; *O*'s are the second numbers.

The EP game matrix is similar in structure to that used by Deutsch (1957) in the study previously described.

FIG. 1. The game matrices for four power conditions.

from four to eight *S*s were randomly assigned to one of the four power conditions. Each *S* within a particular group was then exposed (by a random procedure) to one of three game strategies from his partner.

Procedure

Laboratory setting. Once in the laboratory, the *S*s were placed in cubicles so that they were unable to determine the identity of their game partners. In order to communicate their game choices, the *S*s used an electric signal switch. The two switches on each *S*'s box were wired separately to the *E*'s master control panel, which was hidden from the *S*s. The wiring scheme permitted the *E* to receive information separately from each *S* as well as to send preprogrammed choices to all *S*s simultaneously. At the outset of the experiment, the *S*s were given a standard set of instructions which are outlined below:

> You are going to participate in a study concerned with game playing behavior. You will be playing with another person whom you cannot see or communicate with. One of you will be called Person 1 (*S*), the other of the pair will be Person 2 (*O*). In this game you can either win money or lose money. This money will be *imaginary*, but we want you to feel and act as if it were real money.

Creation of power conditions. Groups of *S*s were randomly assigned to one of four game matrices, each corresponding to a particular power condition (Fig. 1). The power conditions are defined from the point of view of *O*'s position:

O in Absolute Power 1 (AP 1). *O* (Person 2) has absolute power over *S*. *O* can determine whether *S* wins (Column A) or loses (Column B). *S* however cannot affect *O*'s payoffs; *O* wins $30 for any combination of choices. *O* and *S*'s preferences are thus *coordinated*.

O in Absolute Power 2 (AP 2). This condition is the same as AP 1 with one difference; *O* can win more by choosing B ($40) rather than A ($30). If, however, *O* does choose B, *S* must lose. *O* and *S*'s preferences are *divergent*.

O in Partial Power (PP). *O* can determine whether *S* wins or loses. *S* in this condition can, however, exert some control over *O*'s payoffs. By choosing Row Y rather than Row X, *S* forces *O* into a position whereby he can win only $10 or $20, rather than a possible $30 or $40. This condition differs from AP 2 only with respect to the fact that *S* has retaliatory power in the Partial Power Condition.

O in Equal Power (EP). Both players are equally powerful in terms of the extent to which each player can determine the gains or the losses of the other. Each player can achieve his maximal gain by a choice which is detrimental to the interests of the other. (Row Y for *S* and Column B for *O*.) If a YB combination does occur then both players lose. The EP condition differs from the PP condition for *S* in two important respects: (*a*) *S* can exert a greater control over *O*'s payoffs and; (*b*) *S* has greater incentive to choose Y since he can maximize his gain ($40). Under PP a choice of X would enable *S* to maximize his gain ($30).

Game playing instructions. Regardless of the particular matrix utilized, the social implications of any combination of choices were explained to the *S*s in the following manner (illustration is based upon the EP condition):

> Person 1 chooses first between Row X or Y for each trial. Person 2 knowing Person 1's choice then makes his choice of Column A or B. (Before each of you makes his choice you will answer two questions on your game score form.) If Person 1 were to choose X he could either win $30 or $40 depending upon what Person 2 decides to choose. If Person 2 chooses A, both players win $30 whereas if he chooses B, Person 2 would win $40 while Person 1 loses $30. If Person 1 chose Y he could win $40 or lose $20. Here if Person 2 chose A he would lose $30 (Person 1 wins $40) whereas, if he chose B, both would lose $20.

All groups were given the same instructions regarding their goal in playing the game. They were instructed as follows:

> This is the orientation, or the attitude you should

have in playing this game. *If you are Person 1*(S), *your motivation should be to win as much money as you can for yourself.* You are not interested in what happens to the other person. You are not out to help the other person or to beat him. You are merely out to win as much as you can and lose as little as possible and you should feel it's important for you to do so. *If you are Person 2* (O), *you can decide for yourself just how and for what purpose you want to play the game.*

A basic requirement of the experimental situation was that the *S*s understand the mechanics of the game playing procedure as well as the interpersonal effects of their game choices. A twelve-item test was administered to the *S*s just before the game to check on the adequacy of their game understanding. Any *S* who made more than one error on this test was dropped from the experimental sample. Only nine *S*s (4%) failed by this criterion.

Creation of three game strategies. The *E* fed back a preprogrammed series of choices to the *S*s in his role as Person 2. Before describing the game strategies, it is necessary to list and define the basic types of game choices and their social implications:

Cooperative Choice (*coop. choice*) refers to a choice of X or A which serves to maximize the mutual gain.

Noncooperative Choice (*noncoop. choice*) refers to a choice of Y or B which serves to maximize the recipient's loss and minimizes the combined gain of the dyad.

Trusting Behavior refers to a coop. choice based upon expectation of receiving a coop. choice from the other player (*O*).

Suspicious Behavior refers to a noncoop. choice based upon an expectation of receiving a noncoop. choice from *O*.

Exploitative Behavior refers to a noncoop. choice based upon an expectation of receiving a coop. choice from *O*. (EP condition only)

Risk-Taking Behavior refers to a coop. choice based upon an expectation of receiving a noncoop. choice from *O*.

Each *S* received one of three types of game strategies from *O*:

1. *An Unconditionally Cooperative Strategy: O* makes a coop. choice (A) on every trial regardless of what *S* chooses.

2. *A Conditionally Cooperative Strategy:* On the first trial *O* makes a coop. choice. However, for each trial thereafter *O*'s choice is *reciprocal* to that of *S*. *O* makes a coop. choice only when he receives one from *S*.

3. *A Noncooperative Strategy: O* makes a noncoop. choice (B) for every trial regardless of what *S* chooses.

After the data were analyzed, it was discovered that because some *S*s consistently made a cooperative offering in the face of a prescheduled cooperative strategy, they had the experience of mutual cooperation for six trials. These *S*s, therefore, satisfied the requirements for assignment to both the Conditional and the Unconditionally Cooperative Strategy. Moreover, they represented a select group in terms of personality predispositions as well as game experience. For these reasons this group was partialled out for separate analysis and labeled the *Residual Cooperative Group*.

Reverse trial and postgame questionnaire. After six trials of the game, *S*s were informed that for the seventh trial their roles would be reversed. *S* would now choose last. *E* "fed" *S*s in all conditions a coop. choice for the Reverse Trial. At the close of the game, *S*s were asked to complete a postgame questionnaire that tapped their attitudes toward *O* and interpretations of his game playing patterns.

Informal interviews with a number of the *S*s after they played the game revealed that they became quite involved in the game interaction. Several demanded to know the identity of their partners whom they felt had "exploited" or "double-crossed" them.

TABLE 1

PERCENTAGE OF *S*s WITH TRUSTING AND SUSPICIOUS CHOICES IN EACH POWER CONDITION

(Trial 1)

	Absolute Power 1	Absolute Power 2	Partial Power	Equal Power
Trust[a]	43.1	22.0	30.9	24.2
Suspicion[b]	27.6	39.0	32.4	45.2
Residual[c]	29.3	39.0	36.7	30.7
	100%	100%	100%	100%
	$N = 58$	$N = 59$	$N = 68$	$N = 62$
		χ^2	df	p (1 tail)
Hypothesis 1	$AP_1 > AP_2$	5.92	2	$<.03$
Hypothesis 2	$AP_2 > PP$	—	2	*n.s*
	PP vs. EP	—	2	*n.s*

[a] *S* expects a coop. choice from *O* and makes a coop. choice.

[b] *S* expects a noncoop. choice and makes a noncoop. choice.

[c] Residual category includes both of the following:

(*a*) *S* expects a coop. choice and makes a noncoop. choice (exploitative for the EP condition).

(*b*) *S* expects a noncoop. choice and makes a coop. choice (risk taking).

RESULTS AND DISCUSSION

Hypothesis 1. The greater the gain *O* can derive from noncooperative as compared to cooperative behavior, the more likely is *S* to behave suspiciously.

Hypothesis 2. The greater the power *O* has relative to *S*, the more likely is *S* to behave suspiciously.

In order to capture the effects of differences in power relationships in their purest form, the percentage of *S*s who were trusting or suspicious at the outset of the game (Trial 1) were compared for the four power conditions. These data are presented in Table 1.

Hypothesis 1 receives support by virtue of the fact that a significantly greater proportion of *S*s act suspiciously when initiating an interaction with an Absolutely Powerful *O* who stood to gain from a noncoop. choice as

compared to one who did not ($AP_2 > AP_1$). This prediction is also supported by the finding that *S*s make a significantly greater mean number of suspicious choices during the game (Trials 2–6) in the AP_2 as compared to the AP_1 condition (see Table 2).

The test of Hypothesis 2 was based upon a comparison of *S*'s behavior when confronted with an *O* who has Absolute Power (AP_2) as compared to where *O* has Partial Power and *S* could retaliate. Although Table 1 indicates that the differences in the percentage of suspicious choices for Trial 1 were in the predicted direction, they did not attain statistical significance. Table 2, however, shows that the *S*s made a significantly lower mean number of suspicious choices in a relationship where they could retaliate (PP) as compared to one where they could not (AP_2). These findings (Trials 2–6) taken together with the first trial data substantiate Hypothesis 2.

Of interest is a comparison of the Equal Power with the Partial Power Condition for Trial 1. The EP conditions differ from the PP condition in two respects: although *S* can exert greater control over *O*'s outcome this is accompanied by a greater incentive to make a noncoop. choice (see Fig. 1). The absence of a reliable difference between the EP and PP conditions suggests that an increase in *S*'s relative power does not necessarily increase the probability of trusting behavior if the heightened power is accompanied by an inducement to be noncooperative. (See Table 1.)

Hypothesis 3. Under the conditions where *S* receives a Cooperative as compared with a Noncooperative Strategy: (*a*) *S* is more likely to trust *O*; (*b*) *S* tends to develop a greater liking for *O*.

Hypotheses 3a and 3b were tested by applying an analysis of variance to test the effects of four power conditions and four game strategies upon the development of trust and sentiment. With respect to the mean number of trusting choices made during the game (Trials 2–6) the only significant *F* ratio was derived from the main effects of game strategies ($p < .01$). Postmortem *t* tests (Snedecor, 1956) support Hypothesis 3a in that any type of Cooperative Strategy employed by *O* induced a significantly greater number of trusting choices than the Noncooperative Strategy. There was no reliable difference between the Conditional and Unconditionally Cooperative Strategy.

TABLE 2
MEAN FREQUENCY OF TRUSTING AND SUSPICIOUS CHOICES FOR THE COOPERATIVE STRATEGIES UNDER THREE POWER CONDITIONS
(Trials 2–6)

	Means			*t*	
	AP_1	AP_2	PP	AP_1 vs. AP_2	AP_2 vs. PP
Trust	2.80	2.26	2.83	1.33*	1.31*
Suspicion	0.97	1.50	0.85	1.75*	2.20*

* $p < .05$, one-tailed test.

The index of sentiment was based upon the mean of *S*'s responses to nine evaluative items of the Osgood Semantic Differential (Osgood, Suci, & Tannenbaum, 1957). The scores ranged from 1 (dislike) to 7 (like). An analysis of variance of the mean sentiment score for the 16 experimental groups yielded a significant *F* ratio for the main effects of game strategies ($F = 43.7$, $p < .001$) as well as a significant interaction effect. (Power × Strategies, $F = 2.1$, $p < .01$) Postmortem *t* tests indicated, as predicted by Hypothesis 3b, that any type of Cooperative as compared to Noncooperative Strategy induced a greater liking for *O*.

An inspection of the 16 cells suggested that the interaction effect was a function of the reversals in the direction of the differences between the two types of Cooperative Strategies. Under the EP condition the Conditional Cooperative group had a higher mean sentiment score than the Unconditionally Cooperative group; however, under the Unequal Power Conditions (AP_1, AP_2, PP) this difference reversed itself. A *t* test of the mean difference between these two pairs of means yielded a significant interaction ($t = 2.21$, $df = 41$, $p < .02$). Thus, a Conditionally Cooperative *O* was evaluated more favorably than an Unconditionally Cooperative *O* if *S* had Equal Power; however, the Unconditionally Cooperative *O* was evaluated more favorably under conditions where *S* was *dependent* upon *O*.

A behavioral index of sentiment was available through the fact that on the final trial (#7) *S* and *O* reversed roles so that *O*

TABLE 3
MEAN NUMBER OF COOPERATIVE CHOICES FOR THREE GAME STRATEGIES
(Equal Power Condition)

Strategies	Means	Groups Compared	t test
	Cooperative Choices		p(2 tailed)
Cond. Coop.	3.47	Cond. vs. Uncond. Coop.	$<.01$
Uncond. Coop.	1.43	Cond. vs. Noncoop.	$<.01$
Noncoop.	1.00	Uncond. vs. Noncoop.	*ns*

chose first with a programmed coop. choice and S chose last. In this situation the Ss (under conditions AP_2, PP, and EP) were provided with an opportunity of gaining an additional \$10 at a cost of \$30 to O (noncoop. choice) or of accepting a lesser gain (−\$10) thereby enabling O to win \$30 (coop. choice). In effect, a coop. choice at this point represents an abandonment of a purely individualistic orientation. The results, in terms of the proportion of coop. choices made under each strategy are as follows: Residual Cooperative, 88%; Conditional Cooperative, 52%; Unconditional Cooperative, 39%; and Noncooperative, 21%. In line with Hypothesis 3 each Cooperative Strategy differed significantly from the Noncooperative Strategy ($p < .05$; cells corrected for unequal n). There were no significant differences between the Conditional and the Unconditional Cooperative Strategy.

Hypothesis 4. Under conditions where S and O are in an Equal Power relationship and can achieve their respective maximal gain via noncooperative behavior, the application of a Conditional rather than an Unconditionally Cooperative Strategy is more effective in inducing S to play cooperatively. Hypothesis 4 presumes that S learns that O's cooperation is not contingent upon his own cooperation when playing with an Unconditionally Cooperative O, whereas in playing with a Conditionally Cooperative O he learns that his own cooperation is necessary in order to elicit O's cooperation.

The data in Table 3 corroborates Hypothesis 4. Ss chose a significantly greater number of coop. choices when playing with a Conditional

TABLE 4
THE RATIO OF EXPLOITATIVE/NONCOOPERATIVE CHOICES FOR THREE MOTIVATIONAL STRATEGIES
(Equal Power Condition)

Exploitative All-Noncooperative[a]	Noncooperative	Unconditional Cooperative	Conditional Cooperative
Ratio = 0	20	7	15
Ratio > 0	2	7	4

p Value[b]
Uncond. vs. Cond. Coop. $<.01$
Uncond. vs. Noncoop. $<.01$
Cond. vs. Noncoop. *ns*

[a] The ratio of exploitative/noncoop. choices is based upon the proportion of $\frac{\text{Y choices and A expectation}}{\text{Y choice with A or B expectation}}$ for each S. The distributions were highly skewed, thus a dichotomy at O and $>O$ was made.

[b] Fisher exact test.

as compared to an Unconditionally Cooperative O. There was no significant difference, however, between the mean number of coop. choices between the Unconditionally Cooperative and the Noncooperative Strategy. By teasing out those noncoop. choices which were based upon an expectation of cooperation (exploitation) from those based upon an expectation of noncooperation (suspicion) it was possible to make a precise analysis of S's intentions.

Table 4 presents the proportion of noncoop. choices based upon an exploitative intention for the three types of game strategies. The results indicated that the S's noncoop. choices are more likely to be based upon an exploitative intention when playing with an Unconditionally Cooperative O as compared with a Noncooperative or Conditionally Cooperative O. Apparently S was quick to learn that he could satisfy his individualistic goals most efficiently by exploiting O's rigid pattern of cooperation. In spite of the fact that Ss in this condition won more money they did not show as great a liking for the Unconditional as compared to the Conditionally Cooperative O. How can this finding, which on the surface contradicts reinforcement theory as well as the notion of cognitive balance be explained?

S is confronted with a pair of inconsistent

cognitions; *O* (who is presumed to like himself) consistently rewards *S*'s exploitative behavior. *O*, in effect, acts in a "self-defeating" manner. *S* can resolve the cognitive imbalance in one of three basic ways: by disengaging *O* from any responsibility for his self-defeating behavior; by reinterpreting the meaning of the loss for *O*; or by inverting the tacit assumption that *O* likes himself to *O* dislikes himself. The data on motive attribution serves to shed some light on the nature of *S*'s cognitive reorganization.

S's motive attribution. Data pertaining to the type of motive *S* attributes to *O* are derived from *S*'s response to Item 3 of the Postgame Questionnaire which asked: "What do you think *O*'s purpose or goal was in playing the game in this manner?" *S*s could select from one of four fixed alternatives or choose to write in their own response. (11% of the *S*s chose the open-ended alternative.) All of the *S*s' responses were classified into one of the three following categories: (*a*) a *Cooperative Motive* defined as *O* seeks to maximize the mutual gain, (*b*) an *Individualistic Motive* defined as *O* seeks to maximize his own gain and/or *O* seeks to outdo *S*, and (*c*) an *Out of the Field Motive* defined as *O* does not wish to win money or *O* does not understand the game. Table 5 presents the proportion of *S*s who attribute each of these three motives for two types of Cooperative Strategies and two types of Power Conditions. It shows that where *S* has equal power the Unconditionally Cooperative Strategy is more likely to induce the attribution of an Out of the Field Motive, whereas, under conditions where *S* is dependent upon *O*, this strategy is more likely to induce *S* to attribute a Cooperative Motive.

Thus, an *S* who interacts with an Unconditionally Cooperative *O* (EP) resolves his incompatible impressions of *O* by attributing to *O* a disinterest in the goal or sheer cognitive confusion. In effect, the meaning of the loss for *O* has been reinterpreted, and/or *O* is disconnected from responsibility for his behavior. These findings also serve to parallel and complement the results with respect to *S*'s sentiment toward *O*. On theoretical grounds we would expect *S* to have a greater liking for an *O* whose coop. choices stem from an intrinsic desire for group gain than for an *O* whose coop. choices are accidental or only a derivative of a more primary self-centered purpose.

TABLE 5

THE PERCENTAGE OF *S*s WHO ATTRIBUTE EACH OF THREE MOTIVES TO *O* UNDER TWO POWER CONDITIONS AND TWO GAME STRATEGIES

O's Motive	Equal Power Condition[a]		Unequal Power Condition[b]	
	Conditional Cooperative (N = 20)	Unconditional Cooperative (N = 14)	Conditional Cooperative (N = 43)	Unconditional Cooperative (N = 32)
% Cooperative	75.0	57.1	52.4	100.0
% Individualistic	25.0	7.1	35.7	00.0
% Out of Field	00.0	35.8	11.9	00.0
	100.0%	100.0%	100.0%	100.0%

[a] In Equal Power Condition $\chi^2 = 9.01$; df 2; $p < .02$.[c]

[b] In Unequal Power Condition $\chi^2 = .9$; df 2; $p < .01$.

[c] Although the requirements of expected cell frequencies of 5 for using χ^2 were not met, the p value obtained suggests that the null hypothesis may be rejected.

SUMMARY

A two-person nonzero-sum game was devised to test some hypotheses regarding the influence of differences in social power and the strategy with which it is used upon the development of interpersonal trust. One member of the dyad was an *S* while the other was a programmed experimental confederate (*O*). *O* interacted with *S* in one of three conditions of relative power: *O* in Absolute Power, Partial Power, and Equal Power. *O* employed one of three types of game strategies in each power condition; Conditional Cooperation, Unconditional Cooperation, and Noncooperation.

The major findings supported Deutsch's theory of trust. They were the following:

1. *S* is more likely to engage in trusting behavior the greater the power he has relative to *O*.

2. Under conditions of Equal Power, *S* tends to respond to an Unconditionally Cooperative *O* with exploitative game behavior and less liking, whereas he tends to be cooperative and have a greater liking for the Conditionally Cooperative *O*. This difference reversed itself under the Unequal Power Conditions.

These findings are of particular relevance to the problems inherent in the socialization of a child or a patient and they suggest some of the conditions under which exploitative or altruistic motives may develop. More specifically this experiment demonstrates that a Conditionally Cooperative Strategy is most likely to facilitate the transformation of an individualistic to a cooperative orientation, particularly under conditions where there is a high degree of mutual interdependence.

REFERENCES

DEUTSCH, M. Conditions affecting cooperation. *USN tech. Rep.*, 1957, Contract NONR-285 (10).

DEUTSCH, M. Trust and suspicion. *J. conflict Resolut.*, 1958, **4**, 265–279.

DEUTSCH, M., & SOLOMON, L. Reactions to evaluations of others as influenced by self-evaluation. *Sociometry*, 1959, 22, **2**, 93–111.

FARR, J. *The effects of a disliked third person upon the development of mutual trust.* Paper read at APA, New York, September 1957.

FESTINGER, L. *A theory of cognitive dissonance.* White Plains: Row, Peterson, 1957.

HEIDER, F. *Psychology of interpersonal relations.* New York: Wiley, 1958.

LOOMIS, J. Communication, the development of trust, and cooperative behavior. *Hum. Relat.*, 1959, **4**, 305–315.

OSGOOD, C., SUCI, G., & TANNENBAUM, P. *The measurement of meaning.* Urbana: Univer. Illinois Press, 1957.

SCHELLING, T. C. The strategy of conflict: Prospectus for a reorientation of game theory. *J. conflict Resolut.* 1958, **3**, 203–264.

SNEDECOR, G. *Statistical methods.* Ames: Iowa State Coll. Press, 1956.

(Received July 15, 1959)

Journal of Abnormal and Social Psychology
1960, Vol. 61, No. 2, 231–238

SOCIAL EFFECTIVENESS AND SYMPTOMATIC BEHAVIORS

EDWARD ZIGLER[1]
Yale University
AND LESLIE PHILLIPS
Worcester State Hospital and Clark University

IN A RECENT study, Zigler and Phillips (1959) investigated the relationship between mental disorder and case history data. Employing the variables of age, intelligence, education, occupation, employment history, and marital status as referrents of adaptive adequacy, they examined the differences between a state hospital population and the normal population of the same geographical region. They also investigated how each of the major groups of functional disorders differed from the normal population and from one another on these variables.

Despite the limitations incurred through the use of a restricted hospital population and the employment of as unreliable an index of deviant behavior as an individual's diagnosis, Zigler and Phillips found that: (*a*) the hospitalized individual tends to be drawn from the more socially impoverished segment of the population, i.e., from among those individuals who have the most difficulty in meeting the tasks society sets; (*b*) the diagnostic groups that were included in the hospital population showed systematic differences on these biographical variables. The schizophrenics and character disorders were similar to each other and were drawn from the most inadequate portion of the population. The manic-depressives and psychoneurotics were similar to each other and were drawn from an intermediary segment of the population, being characterized by a higher level of adaptive adequacy than that exhibited by schizophrenics and character disorders; (*c*) females in the hospital population tended to differ less strikingly from the normal population on these variables than did the males. Females were more likely to fall into the manic-depressive and psychoneurotic than into the schizophrenic and character disorder groups, and in general exhibited a higher level of premorbid social effectiveness than did hospitalized males. The relationship of adaptive adequacy to other variables has been investigated in both normal and pathological populations. It has been found to be related to prognosis in cases of schizophrenia (Phillips, 1953), success in the rehabilitation of schizophrenics (Farina & Webb, 1956), and, in normal subjects, to level of perceptual cognitive development (Lane, 1955).

The finding that a relationship exists between membership in a particular diagnostic group and the level of adaptive adequacy manifested, as measured by the biographical variables, may be related to the position recently advanced by Phillips and Rabinovitch (1958). After finding that certain symptoms tend to occur together or to be mutually exclusive in their appearance, they proposed a developmental hierarchy for categories of symptoms. Basing their proposal on conceptual grounds, Phillips and Rabinovitch (1958) suggested that symptoms indicative of Self-Deprivation and Turning against the Self, reminiscent of the manic-depressive syndrome, implied an introjection of social standards congruent with a relatively adequate level of social maturity. In contrast, symptom clusters indicative of either an Avoidance of Others or Self-Indulgence and Turning against Others, which may be taken to correspond to the schizophrenic and character disorder categories, respectively, were hypothesized as representing developmentally lower levels of psychological function. While this view is supported by the findings of Zigler and Phillips (1959), their data do not support Phillips and Rabinovitch's view that a difference in social adequacy characterizes the last two categories.

From the conceptual position advanced by Phillips and Rabinovitch and the empirical findings of Zigler and Phillips, certain predictions concerning the relationship between

[1] his investigation was conducted while the senior auth°r was at the University of Missouri and the Worcester State Hospital, Worcester, Massachusetts, and was supported by the Dementia Praecox Research Project, Worcester State Hospital, and a research grant (M-896) from the National Institute of Mental Health, United States Public Health Service.

symptomatology and adaptive adequacy may be generated. If the biographical variables enumerated earlier are thought of as indices of adaptive adequacy, i.e., adequacy in meeting societal expectancies, then the following relationships should obtain:

1. Individuals who manifest those symptoms conceptualized by Phillips and Rabinovitch as indicative of Turning against the Self will be more socially effective than individuals who manifest symptoms indicative of either an Avoidance of Others or Self-Indulgence and Turning against Others.
2. Since Zigler and Phillips found that hospitalized females tended to exhibit a higher level of adaptive adequacy than did hospitalized males, females should more frequently manifest those symptoms indicative of Self-Deprivation and Turning against the Self than do males. Males, on the other hand, should more frequently manifest those symptoms indicative of either an Avoidance of Others or Self-Indulgence and Turning against Others than do females.

It was the purpose here to discover if these relationships do indeed occur. The study also investigated differences in social effectiveness between individuals who manifest symptoms indicative of Avoidance of Others and those whose symptoms are indicative of Self-Indulgence and Turning against Others. Such an investigation may throw further light on the disparity between Phillips and Rabinovitch's proposal that individuals whose symptoms are indicative of Self-Indulgence and Turning against Others will evidence a higher level of adaptive adequacy than individuals whose symptoms are indicative of Avoidance of others, and the empirical findings of Zigler and Phillips which failed to provide support for such a formulation.

METHOD

The study was based on an examination of the case history data of 1,053 patients admitted to Worcester State Hospital during a 12-year period (1945–1957). This sample includes all patients referred to the Hospital Psychology Department for personality study during this period. While the sample includes a wide variety of cases, extremely deteriorated individuals, obviously organic cases, and very agitated patients are seldom referred for diagnostic appraisal and are not adequately represented in this sample. The sample was divided into a male (N = 652) and female (N = 401) group.

Categorization of symptoms. A symptom in the present study, as in the Phillips and Rabinovitch study (1958), refers to the description of a patient's behavior by a psychiatrist at the time of initial institutional contact, or the behavioral description presented by referring physicians as the primary reason for hospitalization. The symptoms include specifications, e.g., assault; general behavior patterns, e.g., irresponsible behavior; thoughts, e.g., sexual preoccupations; somatic reactions, e.g., headaches; and general affect states, e.g., being tense. Examination of the 1,053 case records resulted in the tabulation of 48 discrete presenting symptoms. Those symptoms which occurred only rarely were eliminated from further consideration because their low frequency of occurrence made them unsuitable for the type of statistical treatment employed. Consequently, the results in the male group were based on the analyses of 28 symptoms, and in the female sample, on the analysis of 24 symptoms.

A total of 29 symptoms were used. Of these, 24 had been employed by Phillips and Rabinovitch and placed by them into the categories of Avoidance of Others, Self-Indulgence and Turning against Others, and Self-Deprivation and Turning against the Self. In the earlier study, two of the symptoms reported on here, Perversions except homosexuality and Homosexuality, were treated as a single symptom, Perversions including homosexuality. Four symptoms not employed in the Phillips and Rabinovitch study were also included, Maniacal outbursts, Euphoria, Mood swings, and Fears own hostile impulses. The symptom Maniacal outbursts was placed into the Self-Indulgence and Turning against Others category. Since manic behavior has been conceptualized as a reaction against depression, the symptoms Euphoria and Mood swings were placed into the Self-Deprivation and Turning against the Self category. The symptom Fears own hostile impulses indicates an introjection of social standards in keeping with a higher level of social maturity. This symptom was therefore also included in the Self-Deprivation and Turning against the Self category.

Biographical variables. The categories within each variable were conceptualized as representing steps along a social effectiveness continuum. Because of their inherent order, the variables of age, intelligence, education, occupation, and employment history readily lent themselves to such categorization. For the variable of marital status, developmental theory as employed in recent studies (Jordan & Phillips, 1959; Lane, 1955) was utilized in ordering the categories of this variable. The categories of each variable and their order from low to high were the following:

1. Age—24 years and below; 25 to 34; 35 to 44; and 45 years and above.
2. Intelligence—IQs obtained on a standard intelligence test of 70 or less; 71 to 77; 78 to 84; 85 to 115; 116 to 124; and 125 or above.
3. Education—less than grade school; finished grade school; some high school or trade school; finished high school or trade school; some college including business college, art schools, etc.; and finished college or more.
4. Occupation—*The Dictionary of Occupational Titles* (1949) was used in placing each occupation into the categories of unskilled, semi-skilled, skilled (including foreman), service, clerical and sales, and professional and managerial. Since it was felt that placement of the categories of housewife, student, and other on a social effectiveness con-

tinuum would be arbitrary, individuals falling into these categories were not included in the analyses made on this variable.

5. Employment History—usually unemployed; seasonal, fluctuating, frequent shifts, part-time employment; and regularly employed.
6. Marital Status—single; remarried, separated, divorced, and widowed, with or without children; married without children; and married with children.

RESULTS

The first prediction to be tested was that the occurrence of particular symptoms in an individual's case history was related to that individual's social effectiveness as indicated by the biographical variables. A symptom × biographical variable contingency table for each sex was set up for each symptom on each biographical variable, e.g., tense × age, tense × intelligence, etc. Chi square was employed to determine whether the occurrence or nonoccurrence of a symptom in an individual's case history was independent of the category the individual fell into on each of the biographical variables. The .20 significance level was chosen for these tests. This modest significance level was used since the overall test of the prediction rested not upon a single finding but upon the existence of a predicted directionality in the pattern of symptom × biographical variable relationships.

The results of these analyses for men and women are presented in Tables 1 and 2, respectively. A plus indicates that the symptom was manifested more frequently by individuals who fell into the higher categories of the biographical variable, i.e., into the categories indicative of a higher level of social effectiveness. A zero indicates that the symptom was

TABLE 1

RELATIONSHIP OF SYMPTOMS TO LEVEL OF SOCIAL EFFECTIVENESS IN MEN[a]

Symptoms	*N*	Age	Marital Status	Employment History	Occupation	Education	Intelligence
Category 1—Self-Deprivation and Turning against the Self:							
Depressed	181	+	+	+	+		+
Suicidal attempt	69			+	+	+	+
Euphoria	24	+			+	+	+
Fears own hostile impulses	28		+	+		+	+
Insomnia	29		+	+		+	+
Self-depreciatory	57				+	+	+
Mood Swings	27	+				+	
Bodily complaints	86	+					
Suicidal ideas	68			+		0	+
Tense	213		+		0	+	
Headaches	39		+				0
Category 2—Self-Indulgence and Turning against Others:							
Maniacal outbursts	38		+				
Perversions (except homosexuality)	49	+		+		−	0
Drinking	148	+	0	0	−	0	0
Rape	32		+		−	−	
Robbery	66	−	−	0	−	−	0
Irresponsible behavior	33	−		0	−		
Assaultive	88	−	−		−		0
Threatens assault	71			−	−		
Emotional Outbursts	83	−	−	−			
Homosexuality	34	−	−			−	
Category 3—Avoidance of Others:							
Suspiciousness	195	+	+		+		
Perplexed	70					+	
Bizarre ideas	49		−		+		
Hallucinations	104	0	−				
Sexual preoccupations	61	−	−	0			
Apathetic	38		−				−
Withdrawn	92	−	−	−	−		−

[a] + = high level of social effectiveness; 0 = middle level of social effectiveness; − = low level of social effectiveness.

TABLE 2
RELATIONSHIP OF SYMPTOMS TO LEVEL OF SOCIAL EFFECTIVENESS IN WOMEN[a]

Symptoms	N	Age	Marital Status	Employment History	Occupation	Education	Intelligence
Category 1—Self-Deprivation and Turning against the Self:							
Suicidal attempt	69				+	+	+
Suicidal ideas	68					+	+
Euphoria	25			+		+	
Doesn't eat	28	+		+			
Bodily complaints	60	+					
Tense	138	+			0	+	+
Headaches	23	+	0				
Self-depreciatory	59	0	+	+	+	+	0
Depressed	167	+	+		0		0
Fears own hostile impulses	21	−	+				
Mood swings	27	−					
Category 2—Self-Indulgence and Turning against Others:							
Drinking	33	0	0	+			
Assaultive	37	−				−	
Maniacal outbursts	49	−	−	−			
Threatens assault	23	−	−	−	−		
Irresponsible behavior	36		−		−	−	−
Emotional outbursts	60	−	−	−	−	−	
Category 3—Avoidance of Others:							
Perplexed	72	−	+		+	+	0
Hallucinations	68	−		+	+		
Bizarre Ideas	34	0		−	+		
Sexual preoccupations	33	0		0			
Apathetic	38			−	−	0	
Withdrawn	53		−				
Suspiciousness	132		−	.			−

[a] See footnote Table 1.

manifested more frequently by individuals who fell into the middle categories on the biographical variable. A minus indicates that the symptom was manifested more often by individuals who fell into the lower categories of the biographical variables. No mark indicates that no significant relationship was found between the occurrence of a symptom and the categories on the biographical variable. While symptoms were grouped into the original symptom categories employed by Phillips and Rabinovitch, within categories the symptoms were listed from high to low, i.e., from the most pluses to the most minuses. A summary of the findings reported in Tables 1 and 2 is presented in Table 3.

Symptom Category 1 vs. Symptom Categories 2 and 3. As can be seen in Table 3, the findings support the prediction that individuals who manifest those symptoms indicative of Self-deprivation and Turning against the Self are more socially effective than individuals who manifest symptoms indicative of either Avoidance of Others or Self-Indulgence and Turning against Others. For men, the analyses done on symptoms in Category 1 (Turning against the Self) disclosed that in 32 of the 35 instances in which a significant relationship was found between the occurrence of a symptom and a biographical variable, the symptom tended to occur in individuals who fell into the more socially effective categories on the biographical variables. Among symptoms in Categories 2 (Turning against Others) and 3 (Avoidance of Others), in only 10 of 53 instances in which significant relationships were found did the relationships indicate that the symptom tended to occur more frequently in more socially effective individuals. This general pattern was also found in women. For women, the analyses done on symptoms in Category 1 revealed that in 21 of the 29 instances in which a significant relationship was found between the occurrence of a symptom and a biographical

TABLE 3

SUMMARY OF RELATIONSHIPS BETWEEN SYMPTOM CATEGORIES AND SOCIAL EFFECTIVENESS IN BOTH MEN AND WOMEN[a]

Symptom Category	Men			Women		
	No. & % +	No. & % 0	No. & % −	No. & % +	No. & % 0	No. & % −
Category 1 Self-Deprivation and Turning against the Self	32 (91%)	3 (9%)	0 (0%)	21 (72%)	6 (21%)	2 (7%)
Category 2 Self-Indulgence and Turning against Others	5 (14%)	9 (26%)	21 (60%)	1 (5%)	2 (10%)	18 (85%)
Category 3 Avoidance of Others	5 (28%)	2 (11%)	11 (61%)	6 (32%)	5 (26%)	8 (42%)

[a] See footnote Table 1.

variable, the relationships indicated that the symptom tended to occur more frequently in more socially effective individuals. This was true in only 7 of the 40 instances in which significant relationships were found between symptoms and biographical variables in Categories 2 and 3.

Symptom Category 2 vs. Symptom Category 3. As can be seen in Table 3, there was a tendency in the male group for symptoms indicative of an Avoidance of Others to be related to a higher degree of social effectiveness than was true for symptoms indicative of Self-Indulgence and Turning against Others. This tendency was even more marked in the female group.

Category of Symptom × Sex. The second prediction was that females more frequently manifest those symptoms indicative of Self-Deprivation and Turning against the Self than do males, while males more frequently manifest symptoms of either an Avoidance of Others or Self-Indulgence and Turning against Others. A symptom × sex contingency table was set up for each symptom. Chi square was used to estimate whether the occurrence or non-occurrence of a symptom in an individual's case history was independent of that individual's sex. Since this prediction was based on specific symptom × sex relationships, the .05 level of significance was chosen for this group of tests. As can be seen in Table 4, the findings support the prediction made. Of the 7 symptoms in the Self-Deprivation and Turning against the Self category whose occurrence was found to be significantly related to the

TABLE 4

RELATIVE INCIDENCE OF SYMPTOM OCCURRENCE FOR MEN AND WOMEN

Symptom	Occurs More Frequently In
Category 1— Self-Deprivation and Turning against the Self:	
Suicidal attempt	Women
Suicidal ideas	Women
Euphoria	Women
Doesn't eat	Women
Bodily complaints	ns[a]
Tense	ns
Headaches	ns
Self-depreciatory	Women
Depressed	Women
Fears own hostile impulses	ns
Mood swings	Women
Insomnia	ns
Category 2— Self-indulgence and Turning against Others:	
Maniacal outbursts	Women
Perversions (except homosexuality)	Men
Drinking	Men
Rape	Men
Robbery	Men
Irresponsible behavior	Women
Assaultive	Men
Threatens assault	Men
Emotional outbursts	ns
Homosexuality	Men
Category 3— Avoidance of Others:	
Suspiciousness	ns
Perplexed	Women
Bizarre ideas	ns
Hallucinations	ns
Sexual preoccupation	ns
Apathetic	Women
Withdrawn	ns

[a] ns (not significant).

sex variable, all tended to occur more frequently in women than in men. In the other two symptom categories, of the 11 symptoms whose occurrence was found to be significantly related to sex, 7 tended to occur more frequently in men than in women, while 4 tended to occur more frequently in women than in men. It should be noted that it is the symptoms in the Self-Indulgence and Turning against Others category that tend to occur more frequently in men than in women. This finding does not hold for symptoms in the Avoidance of Others category. Of the two symptoms in the latter category whose occurrence was found to be significantly related to sex, both tended to occur more frequently in women.

DISCUSSION

The most striking finding of the present study is that a systematic relationship exists between measures of social effectiveness and particular groups of symptoms. This observation lends considerable support to the view of Phillips and Rabinovitch (1958) that symptoms are indices of an individual's basic orientation towards the self and others and that this orientation is in turn related to the developmental level the individual has attained. The hypothesis that the most socially effective among hospitalized persons manifest symptoms indicative of Self-Deprivation and Turning against the Self is clearly borne out. Individuals manifesting symptoms of either Self-Indulgence and Turning against Others or Avoidance of Others tend to be drawn from a less socially effective segment of the population.

Another deduction from the major hypothesis was that because hospitalized women appear to be drawn from a more socially effective portion of the population than hospitalized men (Zigler & Phillips, 1959), women more frequently manifest higher order symptoms than men. This prediction is supported by the findings. In every case where a significant relationship existed between sex of the individual and occurrence of a symptom in the Self-Deprivation and Turning against the Self category, it was found that the symptom tended to occur more frequently in women. On the other hand, where significant relationships were found between sex and symptoms of the other two categories combined, a majority of the symptoms tended to occur more frequently in men.

There is no support here for Phillips and Rabinovitch's conceptualization that the Self-Indulgence and Turning against Others category is indicative of a higher level of psychological functioning than is the Avoidance of Others category. In fact, the findings suggest that individuals who manifest symptoms in the latter category tend to be more socially effective than those who manifest symptoms in the former category. This tendency of the Avoidance of Others category to be associated with a higher level of social effectiveness than the Turning against Others category is especially marked in women. Further evidence that the former category represents a higher level of functioning than the latter is indirectly available in the finding that in the two instances where a significant relationship existed between sex and the occurrence of symptoms in the Avoidance of Others category, the symptoms tended to occur more frequently in women. Since the highest order symptoms, i.e., those in the Turning against Self category, tended to occur more frequently in women, this suggests that the Avoidance of Others category shares features in common with this category. Contrary to the common relationship between these two categories, symptoms in the Turning against Others category predominantly tended to occur more frequently in men than in women.

The finding that the Avoidance of Others category is associated with a higher level of social effectiveness than the Turning against Others category presents certain theoretical difficulties. To the extent that Avoidance of Others and Turning against Others are related to the schizophrenias and the character disorders, respectively, it follows that schizophrenia is a "higher" level disturbance than the character disorders. This view would run contrary to a generally held position that schizophrenia represents a dominance of what may be genetically considered the most primitive type of functioning, whereas the character disorders represent a somewhat higher level of psychological activity. (Fenichel, 1945; Fried-

man, 1952; Osborne, 1940). This somewhat paradoxical finding challenges explanation.

If, however, one examines those symptoms within the Avoidance of Others category which are regularly associated with the schizophrenic syndrome, such as Withdrawn, it will be observed that these tend to occur in less adequate individuals. It is the symptoms more atypical of this psychotic disorder which occur in more socially adequate patients. If it is then assumed that the Avoidance of Others category is indeed representative of the schizophrenias, the findings of the present study would not demand the assertion that the overt schizophrenic functions more effectively than the individual hospitalized with a character disorder. What has been measured in the present study is the premorbid social effectiveness of individuals. It may be that the schizophrenic actually achieves a higher level of social effectiveness than does the character disorder, but with onset of the illness, he manifests a more encompassing regression to genetically lower levels of function. This view leads to a conception of schizophrenia as a process in which genetically early levels of functioning infringe upon the higher, more recently developed genetic characteristics. On the other hand, the character disorders would be viewed as a failure in reaching higher levels of development.

Finally, social effectiveness has been treated in the present study as if it were a unitary phenomenon. Actually, it would appear more reasonable to view adaptive adequacy as a constellation of a number of relatively independent patterns of behavior. This view is supported by the recent findings (Jordan & Phillips, 1959) that most items on a frequently used social effectiveness scale, the Worcester Scale of Social Attainment (Phillips & Cowitz, 1953), fell into two major clusters. One cluster appeared to represent a factor of social participation; the other was related to striving for personal advancement. In the present study, the biographical items employed to represent the continuum of social effectiveness appear to be more congruent with the "striving" than with the "social participation" dimension. This appears to be true most explicitly for the occupation, education, and employment history variables. The dimension of "social participation," on the other hand, seems to find minimal representation in the case history items employed. Had biographical items more representative of degree of social participation been used, then individuals whose symptoms are indicative of Avoidance of Others might well have scored at a lower level of adaptive adequacy.

The finding that women tend toward self-destructive behaviors, while men manifest either self-indulgence or behaviors injurious to others, is consistent with the earlier finding that hospitalized women tend to be more effectively adaptive in society than hospitalized men (Zigler & Phillips, 1959). The question arises as to why this should be the case. Tentatively, the suggestion can be offered that both social effectiveness and pattern of symptomatic behavior represent the degree to which an individual tends either to resist or to conform to the mores of society. Thus, men appear to be more aggressive and nonconformers. Other studies (Brun-Gulbrandsen, 1958; Child, Potter, & Levine, 1946) have indicated that women are subjected to far greater pressure for social conformity than are men, and that any deviation in their behavior, particularly in the direction of aggression, is looked upon with much greater concern than it is in men.

The present study has dealt with the proposition that symptom clusters are differentially related to the adaptive adequacy of the individual. Two additional propositions have been advanced in this and earlier papers (Phillips & Rabinovitch, 1958). First, the symptom clusters are indicative of one's role orientation in society; second, the symptoms which one manifests, as well as the congruent level of adaptive adequacy, reflect the psychological development or personal maturity which one has achieved. However, the studies so far reported have dealt only with psychopathological populations. It has been established that either Avoidance of Others or Self-Indulgence or Turning against Others characterize socially inadequate individuals, whereas those psychiatric cases who Turn against the Self tend to be drawn from a more effective stratum of society. It is difficult to conceptualize such negatively toned role orientations as typifying normal individuals drawn from equivalent segments

of society. This raises the interesting question of whether consistent patterns of social interaction may be found in a normal population at different levels of social adequacy and whether these orientations are in any way related to those of pathological individuals of the same levels of social effectiveness.

SUMMARY

The present study tested the following predictions:

1. Hospitalized individuals who manifest those symptoms indicative of Turning against the Self have a more socially effective premorbid history than those with symptoms of either Avoidance of Others or Self-Indulgence and Turning against Others.

2. Hospitalized females more frequently manifest symptoms of Self-Deprivation and Turning against the Self than do males. Males, on the other hand, more frequently manifest symptoms of either Avoidance of Others or Self-Indulgence and Turning against Others than do females.

The study also investigated differences in social effectiveness between individuals who manifest symptoms of Avoidance of Others and those whose symptoms are indicative of Self-Indulgence and Turning against Others. The variables of age, intelligence, education, occupation, employment history, and marital status were employed as indices of social effectiveness. The study was based on an examination of the case history data of 1,053 patients (652 men and 401 women).

Both predictions were supported by the findings. The results also suggested that individuals whose symptoms are indicative of Self-Indulgence and Turning against Others tend to be less socially effective than individuals whose symptoms are indicative of an Avoidance of Others. Present findings were discussed in relation to those of earlier studies as were their implications for patterns of social interaction in a normal population.

REFERENCES

BRUN-GULBRANDSEN, S. Kjønnsrolle og ungdomskriminalitet. Oslo: Universitetsforlaget, 1958.

CHILD, I., POTTER, E., & LEVINE, E. Children's textbooks and personality development: An exploration in the social psychology of education. *Psychol. Monogr.*, 1946, **60**(3, Whole No. 279).

FARINA, A., & WEBB, W. Premorbid adjustment and subsequent discharge. *J. nerv. ment. Dis.*, 1956, **124**, 612–613.

FENICHEL, O. *The psychoanalytic theory of neurosis.* New York: Norton, 1945.

FRIEDMAN, H. Perceptual regression in schizophrenia: An hypothesis suggested by the use of the Rorschach test. *J. genet. Psychol.*, 1952, **81**, 63–98.

JORDAN, E., & PHILLIPS, L. Dimensions of social maturity. Unpublished manuscript, Worcester State Hospital, 1959.

LANE, J. Social effectiveness and developmental level. *J. Pers.*, 1955, **23**, 274–284.

OSBORNE, R. Palaeophrenia: A re-evaluation of the concept of schizophrenia. *J. ment. Sci.*, 1940, **86**, 1078–1085.

PHILLIPS, L. Case history data and prognosis in schizophrenia. *J. nerv. ment. Dis.*, 1953, **117**, 515–525.

PHILLIPS, L., & COWITZ, B. Social attainment and reactions to stress. *J. Pers.*, 1953, **22**, 270–283.

PHILLIPS, L., & RABINOVITCH, M. Social role and patterns of symptomatic behavior. *J. abnorm. soc. Psychol.*, 1958, **57**, 181–186.

UNITED STATES GOVERNMENT PRINTING OFFICE. *The dictionary of occupational titles.* Washington: USGPO, 1949.

ZIGLER, E., & PHILLIPS, L. Case history data and psychiatric diagnosis. Unpublished manuscript, Worcester State Hospital, 1959.

(Received July 28, 1959)

Journal of Abnormal and Social Psychology
1960, Vol. 61, No. 2, 239-245

ROLE ASSIGNMENT AND INTERPERSONAL INFLUENCE[1]

IVAN D. STEINER AND WILLIAM L. FIELD
University of Illinois

LABORATORY investigations of small groups have usually involved the study of contrived groups consisting of strangers who are assembled for a brief period of time and who operate with full knowledge that their associations with one another will not persist into the future. Such contrived groups are manifestly unlike the real-life groups to which we often like to generalize our research findings. In addition to lacking a history and a future, they ordinarily lack a role system that is appropriate to the task at hand. Members are not assigned, nor need they assume, specialized functions; expectations that specific members have special abilities, loyalties, obligations, or responsibilities are seldom created, and the contrived group functions as an unstructured aggregation of individuals.

When investigators have assigned roles to the individuals studied in laboratory experiments, the consequences of such assignments have been very apparent. Carter, Haythorn, Shriver, and Lanzetta (1950) found that assignment of the leadership role to one of the members of a group resulted in leader-follower interactions that were different from those which occurred when the leader was permitted to "emerge" without assignment. Smith (1956) demonstrated that the effect of a silent partner on the productivity and satisfaction of group members depended upon whether or not the group had a role system in which the partner's silence was appropriate. Berlo (1956) assigned each member of his experimental groups the responsibility for performing specific procedural functions, and found that members of such groups were more productive and more satisfied with their performance than were members of similar groups in which no role assignments were made. Borgatta and Bales (1953) have reported that groups in a role playing situation produced interaction profiles which were different from those produced by groups operating without assigned roles. And Steiner and Dodge (1956) discovered that accurate interpersonal perceptions were responsible for increased individual and group productivity when members had not been assigned roles; when the group operated according to a role system, accuracy of interpersonal perceptions had almost no bearing upon individual and group productivity.

The study reported by this paper is like the Steiner and Dodge investigation in that it is concerned with the indirect, rather than the direct, consequences of role assignment. It explores the conjunctive or interactional effects that occur when role assignments are made, and it attempts to demonstrate that persons' perceptions of one another, and their reactions to one another, are contingent upon the presence of a role system. It is hoped that this line of investigation may provide insight concerning the probable generality of conclusions generated by laboratory studies of conformity, attitude change, and group dynamics.

THEORY AND HYPOTHESES

Roles are sometimes "imposed from above" with a minimum of concern for the preferences of the persons involved. Role assignments in military organizations often have this arbitrary quality. Somewhat less arbitrary are the role assignments that occur when persons are appointed or elected to "offices" by their associates; here it is reasonable to presume that the associates have given some consideration to the attitudes and personality of the individual selected. Still less arbitrary are those role assignments that occur when role systems develop slowly, and persons have considerable freedom to "move into" roles that harmonize with their attitudes and personalities. The gradually emerging role structure of a growing business concern is likely to permit this type of self-assignment.

[1] This investigation was supported by a research grant (M-1774) from the National Institute of Mental Health of the National Institutes of Health, Public Health Service (F. E. Fiedler, Principal Investigator). The writers are indebted to Paul Pollock for a superb performance as the experimenter's accomplice, and to William McGuire and Alan Bass for criticisms of the original manuscript.

Two postulates that are concerned with these three methods of role assignment, and with the condition in which no role assignments at all are made, underlie the hypotheses tested by this research.

Postulate 1. The propensity to interpret behaviors as indicative of personal preferences is greatest when no roles are assigned, less when there is self-assignment to roles, still less when assignment is by appointment or election, and least when roles are imposed from above.

The fact that individuals are aware of the discrepancy between personal preferences and the behaviors required by roles imposed from above is indicated by comments such as: "It's not his fault; he's just doing what he has to do," or "The Captain is pretty GI when he's on duty, but he's probably a good Joe when he has a chance." The frequency with which such statements are made by military personnel, employees of large business firms, and others who operate within role-organized agencies suggests that people regularly "make allowances" for role restrictions when estimating others' preferences, wishes, or personality. Postulate 1 asserts that this tendency to make allowances is greater when roles are assigned than when they are not, and greater when roles are assigned arbitrarily than when there is a high degree of self-assignment.

The second postulate contends that the person becomes aware of his tendency, and the tendency of his associates, to make allowances. Consequently, he does not expect that his associates will construe his role-determined behaviors as indicating his personal preferences. Thus, the "civilian soldier" may commit numerous acts of brutality during wartime without feeling that important reference persons will regard him as a brutal person. However, the individual's belief that his behaviors will not be interpreted as indicating personal preferences is undoubtedly most appropriate and most easily maintained when roles are imposed from above; it should be less easily maintained when role assignments are made by election or self-selection.

Postulate 2. The propensity to expect others to construe one's behavior as indicating personal preferences is greatest when no roles are assigned, less when there is self-assignment to roles, still less when assignment is by appointment or election, and least when roles are imposed from above.

Although the postulates deal with four degrees or positions along a presumed continuum, the research to be reported below is concerned with only the two most extreme conditions: no role-assignment vs. role assignment from above. Hypotheses are thus stated in terms which deal only with these two extreme conditions, even though relationships involving the less extreme conditions are clearly implied.

Because the individual is presumed (Postulate 1) to accept observable behaviors as trustworthy clues to an associate's preferences when no roles are assigned, but not when roles are assigned from above:

Hypothesis 1. When no role assignments are made, persons feel more confident of their ability to evaluate the personal preferences of an individual who produces deviant behavior than is the case when roles are imposed from above.

Hypothesis 2. When an imposed role makes it appropriate for a person to produce unpopular and deviant behavior, he is less severely rejected for such behavior than is the case when he produces such behavior without the role having been imposed.

Hypothesis 3. When an imposed role makes it appropriate for a person to produce popular behavior, performance of such behavior leads to a smaller increase in his popularity than is the case when he performs the behavior without the role having been imposed.

According to Postulate 2, an imposed role should decrease the person's concern about whether or not his behaviors will lead to rejection. Consequently, he need not be so alert to observe the reactions that his behaviors are evoking in others; he can afford to persist in a course of action without permitting or inviting others to react to what he is doing. If the activities of the group involve the discussion of a controversial issue, he can present his arguments at great length without allowing others to react to them. At the same time, his associates will realize that these views do not necessarily represent his true feelings, and will feel less need to interrupt in order to correct his "erroneous thoughts." Consequently, if a communication be defined as any continuous

speech behavior produced by a single person, we may hypothesize:

Hypothesis 4. When roles are imposed, members of a discussion group produce longer communications than is the case when no roles are imposed.

Because imposed roles make it easy for the person to conclude that the behaviors of his associates do not represent their true preferences or opinions, those behaviors should have little impact upon his own preferences and opinions. "They don't even believe what they are saying, so why should I?" Consequently:

Hypothesis 5. When an imposed role makes it appropriate for a person to express deviant opinions, his behavior has less effect on the opinions of his associates than is the case when those deviant opinions are expressed without a role having been imposed.

METHOD

Preliminary Session

Ninety-five male college students enrolled in the elementary psychology course at the University of Illinois participated in this study. They received laboratory credit for their participation. During the first session they were assembled in a large classroom and told that they were to serve as subjects (*Ss*) in a study of word meanings. *Ss* were given a three-page instrument on which they were asked to rate the meanings of three concepts: mental illness, desegregation of public schools, and going steady. Each page presented the same ten graphic rating scales, the ends of which were identified by polar adjectives such as wise-foolish, worthless-valuable, and good-bad. Both the seven-step rating scales and the polar adjectives were taken from Osgood and Suci's semantic differential (1955). As is customary with this kind of instrument, *Ss* indicated their appraisals of the three concepts by placing check marks on the scales. Only their appraisals of desegregation of public schools were scored; ratings of the other concepts were obtained to lessen the possibility of sensitizing the *Ss* to the desegregation issue. Furthermore, responses to only six of the ten rating scales were considered in computing the *S*'s attitude toward desegregation of schools; these were six scales that Osgood had found to load heavily on the "evaluative" dimension. Four other scales dealt with potency and activity, and were not scored. Responses at the favorable end of a scale were assigned a value of seven, while responses at the unfavorable end were assigned a value of one. With six items, scores could range from 6 to 42; the obtained scores ranged from 7 to 42, with a mean of 32.5.

Laboratory Session

Subjects. Sixty-eight of the *Ss* who participated in the preliminary session and whose scores on the rating scales indicated that they held nonsegregationist attitudes were selected to participate in the next phase of the research. Their scores ranged from 24 (the theoretically neutral score) to 42 (the most prodesegregationist score possible). Eighty-one of the original 95 *Ss* had scores of 24 or larger, but 3 were discarded because they were Negroes, and scheduling difficulties prevented the use of 10 others.

Each of the 68 *Ss* was paired with another *S* whose attitude score was approximately the same as his own. The two *Ss* in each pair were scheduled to arrive at the laboratory at the same time. Upon arrival the pair was joined by another student who was serving as an accomplice of the experimenter (*E*). The same individual served as accomplice in all of the 34 experimental groups.

The task. The accomplice and the two naive *Ss* were seated at a table, the accomplice always occupying the position at *E*'s left, while one naive *S* sat opposite *E*, and the other sat at *E*'s right. A cardboard sign bearing the letter A was placed in front of the accomplice; a sign bearing the letter B was placed in front of the naive *S* who sat opposite *E*, and a sign bearing the letter C was placed in front of the *S* at *E*'s right. These letters substituted for names throughout the experimental session.

E explained to the *Ss* that their task was to discuss the desirability of desegregation of public schools, and that they should attempt to reach agreement among themselves during the 15-min. period allowed for discussion. *E* then said:

> In order that your discussion will not be one-sided, you are urged to take account of the points of view held by a typical Southern segregationist; a typical Northern minister, priest, or rabbi; and a typical member of the National Association for the Advancement of Colored People.

This admonition was also written on paper and placed before each *S* where it could serve as a reminder during the discussion.

Role assignments. At this point, *Ss* in half of the 34 groups received role assignments. *E* said:

> Now, so that we will be sure that each of these three points of view is adequately presented in your discussion, I'm going to assign each of you the responsibility for presenting one of these three points of view.

E placed a sign in front of Person A (the accomplice) which said Southern Segregationist; signs placed in front of Persons B and C said Northern Clergyman and NAACP Member, respectively. *E* then instructed each *S* to pretend that he was the kind of person designated by his sign, and to advance the views such a person would advance.[2]

[2] The role of Northern Clergyman and NAACP Member were selected for use in this study because it was expected that *Ss* would conceive such roles to harmonize with their own nonsegregationist attitudes, and would not feel that their role restricted their freedom to express their own opinions. Apparently this expectation was correct. *Ss* sometimes prefaced their initial comments by saying, "Because I'm a Northern Clergyman (or NAACP Member) I think desegregation

Groups selected to receive the role-assignment instructions were matched with groups not receiving such instructions on the basis of attitude scores obtained from the preliminary session. *S*s receiving the two types of treatment had identical mean scores of 33.9; the standard deviations of the scores were 4.6 for the role-assigned groups and 4.0 for the no-role groups.

Behavior of the accomplice. The accomplice was always assigned the role of the segregationist in the role groups, and when no roles were assigned he expressed the same segregationist views. He was coached before the experimental sessions began to maintain the position that, although Negroes might not be inherently inferior, their inferior social and economic backgrounds made desegregation unwise during any time in the foreseeable future. Pretest groups were run to give the accomplice practice in presenting standard arguments, and to enable the accomplice and *E* to anticipate the arguments which would be advanced by naive *S*s during the laboratory sessions. In this way the accomplice was equipped with a standard reply for most of the statements made by the naive *S*s. Unanticipated arguments were either ignored, or were answered by the accomplice in a way which seemed consistent with his prescribed segregationist viewpoint.

Certain other restrictions were placed upon the accomplice. He was to attempt to speak first, since the pretests had indicated that it was easier for him to outline his general segregationist viewpoint if he spoke before others had expressed their contrary views. In all but one of the 34 groups he was successful in his attempt to speak first. Furthermore, he was to avoid speaking more often, or for a longer time, in one type of group than in the other. Data to be presented later indicate that he was also successful in meeting this requirement. Finally, even though his "partners" might become emotional during the discussion, he was forbidden to use sarcastic or derogatory language, or to interrupt them while they were speaking. The investigator as *E* observed that these rules were strictly observed.

Measurements. While the discussion was in progress, *E* employed a polygraph to produce a record of the length and serial position of each *S*'s contributions. In order to identify each speaker, three event markers were employed. Whenever a person spoke, *E* pressed a button that activated the appropriate event marker; the button was kept in the depressed position until the *S* finished speaking or was interrupted by another *S*.

At the conclusion of the discussion period, *S*s expressed their attitudes toward desegregation of schools on the same semantic differential scales as had been employed in the preliminary session. They also made other ratings to be described later in this paper.

Explanation to Ss. Before *S*s were dismissed they were given a brief but truthful explanation of the study. Reasons why the accomplice had taken a segregationist stand were revealed, and *S*s were assured that neither the accomplice nor the *E* was a segregationist or interested in convincing them that they should be. Finally, *S*s were requested not to discuss the study with anyone before the semester ended. So far as can be determined, this request was uniformly honored by the *S*s.

of schools is a good idea." After this initial acknowledgment of the role, *E* was unable to observe any difference between the behaviors of *S*s in the role-assigned and no-role groups.

RESULTS

Hypothesis 1

Immediately after the discussion period, *S*s were asked to rate their partners' attitudes toward desegregation on the same 10-item scale as was used in the preliminary session. When these ratings had been made, the *S*s were asked to indicate their confidence in the ratings they had made of each of their partners by placing a check mark on a seven-step graphic rating scale. *E* subsequently assigned numerical values to these check marks which ranged from 1 for a check mark in the space labeled "sheer guesses" to 7 for a check mark in the space at the other end of the scale labeled "very accurate." The mean of the scores assigned to ratings of the accomplice made by *S*s in the role-assigned groups was 4.3, while the mean for the ratings made by *S*s in the no-role groups was 5.0. A four-celled chi square test, with *S*s divided into a high and a low scoring category, yielded a chi square of 5.22, with a probability of less than .03. In computing this chi square, *S*s were divided as closely as possible at the median score (between scores of four and five), and Yate's correction was applied.

The mean of the scores assigned to ratings of nonaccomplice partners in role-assigned groups was 4.4, while in no-role groups it was 4.6. This difference was not statistically significant.

Hypotheses 2 and 3

Rejection of the accomplice and acceptance of the naive partner were evaluated by five sociometric questions administered at the conclusion of the discussion. *S* was asked which of his two partners he would prefer as a roommate, which would prefer him as a roommate, which had the better understanding of him, which he would choose to accompany him on a week-end camping trip, and with which he would be more inclined to discuss his personal problems. Interrelationships between answers to these five questions were determined by phi coefficients to range from +.28 to +.67. All of the interrelationships were significant at the

.02 level or better. Consequently, the five items were pooled to form a sociometric preference scale yielding scores ranging from zero to five. (The score is the number of times the *S* chose the accomplice rather than his naive partner.) The distribution of scores was highly skewed, with 27 *S*s receiving the score of zero because they had not chosen the accomplice in response to any of the five items. For this reason a four-celled chi square test would have been relatively insensitive, and a six-celled test was employed instead. With scores of five, four, and three collapsed to form one category, and scores of two and one collapsed to form another, the obtained chi square value was 7.08 which, with two degrees of freedom, represents a probability of less than .05.

Although this finding indicates that the discrepancy between the sociometric acceptance accorded the accomplice and the naive *S* is greater in the no-role groups, the data do not tell us whether this outcome represents a confirmation of Hypothesis 2, of Hypothesis 3, or of both. Information on this point was obtained from additional postdiscussion ratings. *S*s were given a series of 18 graphic rating scales on which to rate their two partners. Each scale had six steps, and the ends of each scale were identified by polar adjectives of an evaluative nature. Examples are cooperative–uncooperative, and friendly–unfriendly. Ratings at the good ends of the scales were scored six, and ratings at the bad end were scored one. Numerical values were summed across all 18 scales to obtain the *S*'s esteem for a partner. The mean of these esteem ratings received by naive *S*s was 84.4 in the no-role groups and 80.7 in the role-assigned groups. The difference is in the direction predicted by Hypothesis 3, and is significant at the .10 level by a one-tailed *t* test. The mean of the esteem ratings given to the accomplice was 78.3 in the no-role groups and 78.8 in the role-assigned groups. Although this difference is in the direction predicted by Hypothesis 2, it is extremely small and does not approach statistical significance. Apparently the tendency for *S*s in the no-role groups to show a greater sociometric preference for the naive partner than do *S*s in role-assigned groups reflects their relatively higher esteem for the naive partner, rather than lower esteem for the accomplice. It should be noted, however, that in both types of groups the naive partner received significantly more sociometric choices and significantly higher esteem ratings than did the accomplice.

Hypothesis 4

The polygraph records indicated that the average duration of the communications produced by the naive *S*s in the role-assigned groups was 31.5 sec. In the no-role groups the average duration of communications was 23.8 sec. To test the significance of the difference between types of groups, the mean duration of each *S*'s communications was computed, and the two distributions of means were tested by the Mann-Whitney *U* test. The *U* test was employed for this purpose because the distributions were highly skewed. In accordance with Hypothesis 4, the distribution for the role-assigned *S*s was found to be stochastically larger than the distribution for no-role *S*s, the difference being significant at the .04 level by one-tailed test.

The accomplice, whose job it was to maintain the same pattern of behavior in both types of groups, produced communications in the role-assigned groups that averaged 43.5 sec. in length. In the no-role groups his communications averaged 41.0 sec. in length. In the no-role groups his total speaking time averaged 460 sec. per group, while in the role-assigned groups the average was 478 sec. Neither of these differences even approaches statistical significance. Apparently the accomplice was able to abide by his instructions rather well.

Hypothesis 5

The impact of the accomplice on the attitude of the naive *S*s was measured by the discrepancy between *S*s' pre- and postdiscussion attitude scores. The second of these scores was subtracted from the first. The mean difference was +4.6 for *S*s in the no-role groups, and +.7 for *S*s in the role-assigned groups. This outcome is in the predicted direction (large positive difference between pre- and postdiscussion scores represents large change in the direction advocated by the accomplice), and is significant at the .01 level by a one-tailed *t* test.

DISCUSSION

This investigation has employed a univariate design in which numerous variables have been held constant while one variable (role assignment vs. no role assignment) has been manipulated. Whenever this type of design is used there is the possibility that relationships between the independent and dependent variables are contingent upon the level at which other variables have been held constant. Consequently, it is highly desirable that the present study be repeated, and that systematic changes be made in the levels at which variables are controlled. It is especially important that the identity of the accomplice be varied in order to determine whether his personal characteristics and his conception of his role influence the outcome of the research. Other variables that should be subjected to systematic investigation are: the issue that is discussed; the size and direction of the discrepancies between the positions held by the accomplice and the naive *S*s, and the personal characteristics of the naive *S*s. Until such replications have been performed, it can only be said that the hypotheses have been tested under conditions that appear not to be highly atypical, and that they have received fairly strong support.

Although the data support four of the five hypotheses, *S*s were not found to reject the accomplice more completely in the no-role than in the role-assigned groups. Furthermore, *S*s in no-role groups did not judge the accomplice to be more segregationist in attitude than did *S*s in role-assigned groups; both kinds of *S*s judged him to be a strong segregationist. These findings appear to contradict the line of reasoning on which this research has been based, and make it difficult to understand why *S*s in the no-role groups produced shorter communications, yielded more completely to the segregationist view advocated by the accomplice, and esteemed their naive partner more highly than did *S*s in the role-assigned groups. Perhaps a reasonable explanation is that *S*s in the no-role groups were more certain of their appraisals of the accomplice (Hypothesis 1), and that a difference in certainty of appraisal has much the same effect as does a difference in the content of the appraisal itself. However, if this is true, it is still unclear why *S*s in the no-role groups did not indicate less esteem for the accomplice.

Perhaps part of the difficulty lies in the fact that data were obtained from rating scales. College students are undoubtedly reluctant to rate one another unfavorably, and the instrument with which esteem data were obtained permitted them to avoid doing so. On the other hand, the sociometric data were obtained from forced-choice items, and here the accomplice was rated unfavorably by both types of *S*s, and especially so by the *S*s in no-role groups.

Some of *E*'s observations are also pertinent at this point. Because the accomplice was a very persuasive person, and because the research design required that he behave the same way in both types of groups, his performance in the role-assigned groups appeared to be very genuine, and to reflect his true feelings. Because of the enthusiasm with which he expressed segregationist views, many naive *S*s may have "forgotten" that he had been assigned the role of a segregationist, or may have concluded that his assigned role happened to harmonize with his attitudes. If such was the case, some of the effect of role assignment was undoubtedly nullified by his great ability to enact the role. In this respect we undoubtedly failed to approximate the conditions in real-life groups where the individual who is assigned a role to perform is unlikely to be as seemingly sincere or strenuous an advocate as was our accomplice.

An alternative explanation of the data cited in support of Hypothesis 5 must be considered. Janis and King (1954) have found that *S*s who are assigned roles requiring that they advocate a point of view sometimes shift their attitudes to favor the view they have been induced to advocate. Consequently, the failure of our role-assigned *S*s to yield to the accomplice's arguments might be a result of their having played desegregationist roles, rather than a result of their having doubted the sincerity or genuineness of the accomplice's arguments. If this alternative explanation of our data is correct, *S*s assigned the most extreme desegregationist role (NAACP Member) and *S*s who spoke most often and at greatest length while enacting their assigned role, should show the most

resistance to the accomplice's arguments. None of these derivations is confirmed; the obtained probabilities are greater than .70. It therefore seems appropriate to conclude that the findings cannot be explained as consequences of the naive *S*s' enactment of their own roles.

The present study supplements the earlier investigation by Steiner and Dodge (1956). Whereas the earlier study was concerned with the manipulation of interpersonal perceptions, and with the consequences of such manipulations on sociometric preferences and group productivity, the present study has permitted interpersonal perceptions to vary in response to the assignment of roles, and has explored some presumed consequences of these variations in interpersonal perceptions. As in the case of the earlier study, the presence or absence of role assignments has been found to influence the relationship between social psychological variables. Behavior that evokes one pattern of response in a group operating within the framework of a role system has been found to evoke a quantitatively different pattern of response when no roles are assigned. It seems reasonable to conclude that role assignment "interacts" with other variables to produce effects that are very important both to the student of group dynamics and to the practitioner who would apply the scholar's wisdom outside the walls of the laboratory. The study suggests that very great care must be taken when generalizing the findings of laboratory investigations concerned with conformity, attitude change, or group discussion.

SUMMARY

Ninety-five male college students responded to an instrument designed to evaluate their attitudes toward desegregation of schools. From this number, 68 whose attitude scores reflected nonsegregationist views were selected to participate in a laboratory study. Thirty-four groups, each consisting of two *S*s drawn from this sample and another college student who served as an accomplice of the *E*, discussed the desegregation issue for a period of fifteen minutes. In half of the groups the *S*s were assigned roles; the accomplice was always assigned the role of a Southern segregationist while the other participants were assigned the role of a Northern clergyman and the role of a member of the NAACP. In the other half of the groups, participants were not assigned specific roles but were asked to make certain that the views of these three kinds of persons were presented in their discussion. The accomplice endeavored to express the same segregationist views in all groups.

When specific roles were not assigned, the nonaccomplice *S*s: (*a*) had greater confidence in their appraisals of the accomplice's attitudes, (*b*) indicated greater sociometric preference for one another (rather than for the accomplice) at the conclusion of the discussion period, (*c*) produced shorter communications, and (*d*) yielded more to the segregationist arguments expressed by the accomplice. Contrary to expectation, when no roles were assigned, *S*s did not indicate less esteem for the accomplice than was the case when roles were assigned.

It is concluded that role assignments have an important influence on group members' perceptions of one another. Consequently, role assignments are important determinants of members' responses to one another's behavior.

REFERENCES

BERLO, D. Allocation of procedural responsibilities as a determinant of group productivity and satisfaction. Unpublished doctoral dissertation, Univer. Illinois, 1956.

BORGATTA, E. F., & BALES, R. F. Interaction of individuals in reconstituted groups. *Sociometry*, 1953, **16**, 302–320.

CARTER, L., HAYTHORN, W., SHRIVER, B., & LANZETTA, J. The behavior of leaders and other group members. *J. abnorm. soc. Psychol.*, 1950, **46**, 589–595.

JANIS, I. L., & KING, B. T. The influence of role playing on opinion change. *J. abnorm. soc. Psychol.*, 1954, **49**, 211–218.

OSGOOD, C. E., & SUCI, G. S. Factor analysis of meaning. *J. exp. Psychol.*, 1955, **50**, 325–338.

SMITH, E. E. Effects of threat induced by ambiguous role expectations on defensiveness and productivity in small groups. *Group Process Lab. tech. Rep.*, 1956, No. 1. (Nonr-1147(03), NR 170 226, Univer. Colorado) (Mimeo)

STEINER, I. D., & DODGE, J. S. Interpersonal perception and role structure as determinants of group and individual efficiency. *Hum. Relat.*, 1956, **9**, 467–480.

(Received August 3, 1959)

Journal of Abnormal and Social Psychology
1960, Vol. 61, No. 2, 246-254

VOCABULARY KNOWLEDGE AND USAGE OF SCHIZOPHRENIC SUBJECTS:

A SIX-YEAR FOLLOW-UP[1]

LOUIS J. MORAN[2]
Veterans Administration Hospital, Houston, Texas
DONALD R. GORHAM[3]
Veterans Administration Hospital, Perry Point, Maryland
AND WAYNE H. HOLTZMAN
University of Texas

FOR THE past 30 years vocabulary performance, used as a measure of former intelligence level, has played a prominent role in the assessment of intellectual deterioration. In recent years, however, the use of vocabulary measures in this manner has provoked a barrage of criticisms (see Yates, 1956, for a comprehensive review). Foremost among these criticisms, as they apply to deterioration in schizophrenics, has been the contention that vocabulary performance also is impaired in schizophrenics, and, hence, cannot be used as a so-called "stable baseline" from which to judge deterioration in other abilities.

Impairment of vocabulary performance in schizophrenics, if present, could be manifested as a decline coincident with and a function of the psychotic episode itself, or as a gradual erosion with continued illness and institutionalization, or both. In an earlier investigation (Moran, 1953) the first possibility was examined in a cross-sectional study of 40 matched pairs of schizophrenic and normal subjects (*S*s). It was concluded that "... although the schizophrenic may appear to 'define' words as adequately as the control subject, he is less able than the control subject to attain the 'more difficult solutions' of further differentiating the word meanings, or to use the same words in reasoning or in communicating" (p. 16).

The present study reports the results of a follow-up of these schizophrenic *S*s. Yates (1956) has concluded, after a careful review of available literature, that "... a potent (perhaps the most important) factor in inducing a decline in vocabulary is the length of time a patient is institutionalized" (p. 436). The present study tests this hypothesis by a re-examination of the same schizophrenic *S*s with the same battery of word meaning measures after six additional years of institutionalization.

In addition, the current "social adequacy" of the *S*s was rated in the present study. Thus, it was possible also to evaluate the prognostic utility of word meaning measures by relating test performance in 1952 to rated social adequacy in 1958.

In 1952, 54 schizophrenic *S*s completed the entire test battery and one *S* completed all but the word association test. The samples described in the present report were drawn from this total of 55. During the ensuing 6 years, 35 patients remained hospitalized, 5 died, and 12 were discharged and living in the community. The other 3 *S*s could not be located.

USE AND UNDERSTANDING OF WORD MEANINGS IN 1952 AND 1958

Since the hypothesis to be tested here concerned the effect of prolonged institutionalization upon the schizophrenics' use and understanding of word meanings, the 35 *S*s still hospitalized after six years were of primary interest.[4] A full description of this sample

[1] This study was made possible by a grant from The Hogg Foundation for Mental Health, University of Texas. Grateful appreciation is expressed to Ruth M. Hubbard and other members of the Waco VAH staff for their cooperation in carrying out this study.

[2] On leave from University of Texas.

[3] Formerly at Waco VAH.

[4] It would, of course, be of interest to know whether performance of the discharged group had changed in the six years. Several of these *S*s refused to be retested, however, and others were in remote parts of the coun-

may be found in the earlier report (Moran, 1953). In brief, in 1952, all had been diagnosed as chronic paranoid schizophrenic without known organic disease and with an age range of 24 to 57 years (M = 41.3), and had been institutionalized from 1 to 19 years (M = 4.1). Of the 35 hospitalized patients, 30 were retested with the same battery of word meaning measures used in 1952.[5]

Word Meaning Measures

The test battery consisted of the Wechsler-Bellevue (WB) vocabulary subtest, Form 1, and an additional group of seven special word meaning measures. The seven special tests (untimed) used a common core of 25 very familiar words and were designed to afford measurements, for each word, of the following dimensions of word usage and understanding.

1. *Definition.* The first 10 words (Part I), which were names of things, were scored separately from the last 15 words (Part II), which had no "thing" referent. Each of the 25 definitions was scored for (*a*) adequacy and (*b*) abstractness.

2. *Synonym (Recall).* *S* gave as many synonyms as he could think of for each word, and performance was scored for percentage correct synonyms given.

3. *Synonym (Recognition).* The task was to underline the appropriate synonyms among eight alternates for each word, and was scored for percentage correct synonyms underlined.

4. *Sentence Construction.* The 25 words were grouped into 12 sets, ranging from one to three words each. The task was to write a meaningful single sentence for each set, using all of the words in the set, and was scored for total correct sentences written.

5. *Similarities.* Other combinations of the 25 words were used in a 17-item similarities task. Part I consisted of 7 items using only words with a "thing" referent; the 10 items in Part II used only words without a "thing" referent. Each of the 17 items was scored for (*a*) adequacy and (*b*) abstractness.

6. *Analogies.* An 11-item analogies test, constructed around the basic 25 words, was scored for total correct.

7. *Word Association.* *S* gave his first association (one word) to each of the 25 words. Associations were scored on a 4-point scale for relatedness, i.e., "conceptual distance" of response word from the stimulus word.

The purpose of this special battery of tests was to permit an intensive multidimensional examination of the *S*'s knowledge and usage of a few very familiar words. Thus, for each of the 25 words, a measure was provided for (*a*) ability to define, (*b*) conceptual level of definition, (*c*) "breadth" of understanding (number of other words available to express the same meaning as that of the stimulus word), (*d*) preciseness of understanding (percentage correct synonyms), (*e*) ability to communicate (sentence construction), (*f*) ability to form concepts (similarities), (*g*) conceptual level of concepts (similarities), (*h*) ability to reason symbolically (analogies), and (*i*) associative links.

The test forms and detailed scoring instructions for these measures are available from ADI.[6] Every effort was made to administer the tests in the same manner as in 1952. All protocols (the original 30 of the retest 30) were then shuffled and scored independently by two scorers, using the scoring manual constructed for the original study. Interscorer reliabilities on the different measures ranged from .94 to .98. After scoring reliabilities were computed, all scoring differences were resolved jointly by the two scorers in close accord with the manual.

TEST-RETEST RESULTS

No significant ($p < .05$) mean changes occurred in the six-year period, although variability was somewhat greater on second testing, especially on the adequate definitions task. On the whole, test-retest reliabilities were high, considering the small number of items on most tests and the years intervening between performances. The overall picture is one of remarkable stability in the use and understanding of word meanings.

It has been argued that the scoring system for the WB vocabulary subtest may be insensitive to subtle qualitative changes in definitions which are alleged to be charac-

try. Three *S*s living in the community were retested. Inspection of their test-retest scores revealed very slight changes, with inconsistent direction on different tests in the battery.

[5] Five hospitalized *S*s were not retested solely because they were in remote regions of the country. Thus, their omission suggests no a priori reasons for bias in the retest sample.

[6] A copy of the test forms described in this study, with scoring manual, has been deposited with the ADI. Order Document No. 4355 from the ADI Auxiliary Publication Project, Photoduplication Service, Library of Congress, Washington 25, D.C., remitting in advance $3.75 for microfilm or $2.00 for photocopies. Make checks payable to Chief, Photoduplication Service, Library of Congress.

TABLE 1

COMPARISON OF PERFORMANCE BY HOSPITALIZED SCHIZOPHRENICS ON WORD MEANING MEASURES IN 1952 AND 1958

(No significant mean differences; $n = 30$)[a]

Word Meaning Measure	1952		1958		r[b]
	M	*SD*	*M*	*SD*	
Wechsler-Bellevue Vocabulary subtest, Form 1	21.8	6.86	20.4	7.22	.93*
Other Word Meaning Measures					
Definitions					
Adequate definitions, total	24.2	1.02	22.4	3.93	.25
Abstract definitions, total	13.0	5.49	13.0	5.91	.80*
Adequate definitions (Part I)	9.9	.35	9.7	.88	
Adequate definitions (Part II)	14.3	.96	12.7	3.15	
Abstract definitions (Part I)	6.0	3.05	6.3	2.85	
Abstract definitions (Part II)	7.5	3.33	6.7	3.64	
Synonym (Recall)					
Percentage correct synonyms	51.5	18.01	51.1	18.37	.75*
Synonym (Recognition)					
Percentage correct synonyms	73.5	14.43	72.9	15.37	.76*
Sentence construction					
Adequate sentences	7.9	2.55	8.5	2.64	.31
Similarities					
Adequate, total	9.9	3.78	9.1	4.75	.77*
Abstract, total	7.0	4.23	6.9	4.92	.79*
Adequate (Part I)	5.1	1.39	4.7	1.76	
Adequate (Part II)	4.6	2.58	4.4	3.15	
Abstract (Part I)	3.2	1.94	3.2	2.15	
Abstract (Part II)	3.6	2.66	3.6	2.76	
Percentage adequate scored abstract	65.7	20.19	66.7	25.38	
Analogies					
Number correct	6.4	2.91	6.2	2.83	.73*
Word Association[c]					
Relatedness score	38.8	9.69	37.2	10.36	.68*

[a] Mean differences computed by direct difference t for correlated means. Adopted significance level $p < .05$.

[b] Pearson product-moment correlation between first and second testing.

[c] $n = 29$, one *S* did not complete task.

* $p < .05$; correlation significantly different from zero.

TABLE 2

QUALITATIVE SCORES ON WECHSLER-BELLEVUE VOCABULARY SUBTEST, FORM 1, IN 1952 AND 1958

(No significant mean differences, $n = 30$)[a]

Type of Definition	1952		1958		r[b]
	M	*SD*	*M*	*SD*	
Synonym	11.9	5.53	12.3	5.26	.91*
Use and Description	3.0	1.23	2.8	1.15	.58*
Explanation	2.4	2.02	3.3	2.38	.64*
Inferior	2.6	1.75	1.6	1.36	.68*

[a] Mean differences computed by direct difference t for correlated means. Adopted significance level $p < .05$.

[b] Pearson product-moment correlation between first and second testing.

* $p < .05$.

teristic of schizophrenics. Feifel (1949) applied a 5-point scoring system for quality of definition to the Terman Vocabulary List and demonstrated significant mean differences in quality scores between "abnormal" (organic, manic-depressive, schizophrenic) and normal *S*s matched for Terman vocabulary scored in the standard manner. However, essentially the same qualitative scale, used with the WB vocabulary subtest, Form 1, failed to differentiate schizophrenics from normal *S*s in a matched-pair design with *S*s equated for WB vocabulary scored in the standard manner (Moran, Moran, & Blake, 1952).

To determine whether shifts in qualitative level of definition on the WB vocabulary had occurred in the present sample, all definitions, test and retest, were scored independently by two judges in terms of four qualitative levels: (*a*) synonym, (*b*) use or description, (*c*) explanation, and (*d*) inferior. The full rationale of this scoring system, with specific criteria and examples for each category, may be found in Moran (Moran, et al., 1952). Interscorer agreement on qualitative categories was 96%.

The stability of WB vocabulary performance in this sample of schizophrenics, even when scored for quality of definitions, is evident in Table 2. No significant mean differences occurred at any qualitative level. Again, the high retest reliability of the major category, synonyms, attested to the consistency of vocabulary performance over the six-year period. As a further check, qualitative categories were assigned weights of 3 for synonyms, 2 for use or description, 1 for explanation, and 0 for inferior. The weighted scores on test-retest yielded a direct difference t of .23, $p > .70$.

However, still another possibility must be considered. Although no mean differences on the word meaning measures were statistically significant, it may be observed in Table 1 that 13 of the 19 sample means were lower on second testing, though usually by an extremely small margin. A remote possibility still re-

mained that statistically significant differences in direction of change were obscured in comparisons of mean performance levels. To test this hypothesis, each of the eight word meaning measures was rescored for direction of change. On each test, *S* received a 0 if his second score was lower than the first, 1 if the same, and 2 if higher. Six of the measures evidenced no consistent change in direction. However, this weighting system exposed a significant downward shift in WB vocabulary and in the 25-word definitions test.

On the WB vocabulary test, 19 *S*s made a lower score on second testing, 3 the same score, and 8 a higher score. The difference between the observed distribution and that under the hypothesis of no directional change was significant ($p < .05$) as tested by chi square corrected for discontinuity. On the 25-word definitions test (total adequate), 17 *S*s made a lower score on second testing, 9 the same score, and 4 a higher score. This downward shift in scores was also statistically significant.

To summarize at this point: (*a*) 19 comparisons by means of direct difference *t*'s between raw scores in 1952 and 1958 failed to demonstrate a single significant change in mean level of performance on any measure; (*b*) a comparison of 1952 and 1958 performance on WB vocabulary, scored for quality of definition, showed no significant mean differences; but (*c*) when magnitude of change was ignored and direction of change emphasized, a statistically significant decline in performance was found in WB vocabulary and in the 25-word definitions test.

Obviously, the change in WB vocabulary performance was an extremely subtle one. A clue was sought in the study of "off-quadrant" cases, i.e., those below the median on one testing and above on the other. However, as might be evident from the high test-retest correlation ($r = .93$), this attempt was thwarted by the paucity of cases in the off quadrants (one *S* in each).

Qualitatively, one of the most notable differences between the first and second testing was an increased verbosity. The mean number of words used to define the words in the WB vocabulary test increased almost twofold in 1958 (from 119 to 197). An illustration of this tendency is provided in Table 3, which lists definitions to the word *join* by every *S* in 1952 and 1958.

TABLE 3
DEFINITIONS OF THE WORD *Join* IN 1952 AND 1958

Subject	1952	1958
1	Join a church, army	That's when you shake hands with someone and you become in unity
2	Join together	To join something, some kind of lodge or something
3	Put together	Put together or get together
4	Sign up with some kind of organization	Connect up with something else
5	Tie	Connect
6	Cooperation of people, join an organization or hands to be married	Like join a church, a crowd joined together
7	Join something	I don't know
8	To put together, assemble	To put together
9	Together	Put together
10	Anything put together	Join organization, put anything together
11	Put together	Get with; place together or make connections
12	Together	Put together
13	Join clubs	Means to join together
14	To enter into an organization	To become a member of
15	Baptise; connect	Connect
16	To enlist	To enlist; put two things together
17	Join something together	Like join anything together
18	Together	Connect
19	Unite	Bring together
20	Unite	Like putting two pieces together
21	To unite	Join the church
22	Unite	Put together; don't know the correct definition
23	Connect	To put together
24	Join a church	Like join something together
25	Attach	Draw close together; get together
26	Connected	Going together
27	Join a club	Like join the Red Cross; join together; a group of people joined together
28	Unite	To become affiliated with
29	Put together	To make an alliance with; put together; bring together
30	Together	To put together or to join a church or secret organization

It might be hypothesized that the subtle decline in vocabulary performance noted in these *S*s was a function primarily of lessened facility in communicating their knowledge rather than of loss of knowledge per se. This hypothesis is supported by the increased difficulty these *S*s had in expressing themselves,

as evidenced by the increase in number of words used to define test words, and, qualitatively, by the strained expressions illustrated in Table 3. The words in the 25-word definitions test were selected deliberately because of their common everyday usage. Failure to define some of these words adequately, while defining more difficult WB vocabulary words, also supports the hypothesis.[7] Further, the only decline in performance noted in the entire test battery was specific to the two definitions tests, which placed a premium upon verbal facility.

If the very subtle changes on the WB vocabulary test represented a loss of knowledge or failure of memory, more credit would be lost on words at the upper end of the scale. Recall of the meaning of rarely used words like *moiety* or *traduce*, for example, would be more tenuous than for words like *cedar* or *tint*. On the other hand, if the decline represented a failure to communicate knowledge, credit might more easily be lost at any point along the scale. Thus, a test of the hypothesis that the decline was due to difficulty in communicating knowledge rather than loss of knowledge was afforded by examination of the distribution of the 1952–1958 gains and losses in credit for Words 1 through 42 on the WB vocabulary test.[8]

The frequency of loss or gain in credit points (-1, $-\frac{1}{2}$, 0, $+\frac{1}{2}$, $+1$) was tallied for each word in the WB vocabulary test. The distribution of gains and losses throughout the 42 words was highly similar for each word. On Words 1–21, credit was lost at second testing 91 times and gained 51 times; on Words 22–42, credit was lost 82 times and gained 55 times. The hypothesis of loss in verbal facility as opposed to loss of "knowledge" thus received confirmation.

Whatever the nature of this slight drop in performance on definitions tests, what effect would it have upon a "deterioration index" of the type that employed the discrepancy in performance between vocabulary tests and problem solving tasks? For illustrative purposes, a Deterioration Index (DI) of this type was constructed by dividing WB vocabulary raw score into the summed raw scores on analogies and adequate similarities. The resulting ratio thus served as an index of the discrepancy between "former ability level" (size of vocabulary) and present "mental efficiency" (problem solving ability).

The mean DI for the normal *S*s who served as a matched-pair control group in 1952 (see Moran, 1953) was .899. The mean DI in 1952 for the present sample of schizophrenics was .748. This mean difference of .151 was significant ($p < .001$). Thus, such an index in 1952 would have shown the schizophrenic group to be deteriorated.

An index constructed in the same manner from 1958 scores of the schizophrenic group yielded a mean of .750, almost identical to the .748 of 1952. The correlation between the 1952 DI and 1958 DI was .23, $p > .05$. Thus, although individuals changed in relative position, the group as a whole showed no evidence of deterioration by this operation in the six years.

However, in this instance actual former ability level (at least in 1952) was known, making it unnecessary to use inferences from present vocabulary performance. Therefore, a DI was constructed using the actually observed former level (1952 vocabulary performance) in a ratio with current "mental efficiency" (1958 analogies and similarities). The mean of this DI, .686, was significantly ($p < .001$) lower than the DI of .750 obtained when 1958 vocabulary level was contrasted with 1958 performance on analogies and similarities. Thus, actually observed "deterioration" was undetected by a deterioration index of the traditional type.[9]

[7] It should be noted, however, that on this test the only possible direction of change was downward. The mean on first testing was 24.2, where the maximum possible score was 25. Hence, the regression here may be more statistical than psychological.

[8] Scoring criteria on the WB vocabulary test are progressively more stringent through the list. Thus, greater precision in the choice of words is required of definitions for words at the upper end of the scale. Also, fewer *S*s reach the upper end of the scale, so that an instance of credit loss carries proportionately more weight than an instance at the lower end of the scale. Both of these factors, however, are working against the successful outcome of the hypothesis as tested here.

[9] Correlations of each of the three DIs with age and number of years institutionalized were not statistically significant.

DISCUSSION AND CONCLUSIONS

Viewed collectively, the results of this study demonstrated a remarkable stability of performance by hospitalized chronic schizophrenics on word meaning measures taken six years apart. Performance on the kinds of tasks commonly thought to be sensitive to deterioration, e.g., analogies and similarities, showed no significant change in this period. Paradoxically, the only change noted was in the one measure thought to be relatively insensitive to deterioration—definitions. But even here the decline was so small as to go undetected by some of the usual statistical tests.

The downward shift in WB vocabulary score on second testing, though very slight, occurred in about two thirds of the *Ss*. The uniformity of the change was apparent also from the test-retest correlation of .93, i.e., rank order was maintained closely on first and second testing. Decreased verbal facility, rather than loss of "knowledge," seemed to be a primary reason for the poorer performance on second testing.

Was this slight loss in verbal facility due to increased schizophrenic pathology, to the stultifying effect of six additional years of institutionalization, to the normal process of aging, or to a complex interaction of these and possibly other variables? Unfortunately all of these possibilities are inextricably tied together in the present sample. However, certain lines of evidence may be pursued which might tend to implicate some of these variables more than others.

First, if the decline in verbal facility were due solely to increased schizophrenic pathology, it should be reflected most strongly in those tests most "diagnostic" of schizophrenic thinking. Yet observed mean scores on tests that in 1952 reliably differentiated this same schizophrenic group from matched normal *Ss* changed proportionately less than WB vocabulary observed mean score (ignoring for the moment the fact that these observed mean changes were not statistically significant). Furthermore, a significant number of *Ss* showed a slight drop in WB score, but not on the more "diagnostic" tests. It does not seem plausible that performance on measures known to be sensitive to schizophrenic thinking should be unaffected by increased schizophrenic pathology, whereas performance on a measure known to be highly insensitive in this respect showed a significant decline.

Second, if institutionalization per se led to decreased verbal facility, performance level and years of institutionalization should tend to be correlated. The correlation between years hospitalized and WB vocabulary score in 1952 was .01. A more sensitive test would be the relationship between years institutionalized and amount of *change* in WB vocabulary score in the six year period. This correlation was .03. Since in 1952 this sample had been hospitalized from 1 to 19 years, in 1958 the range was from 6 to 25 years. Of those now hospitalized from 9 to 25 years, nine had a lower score on second testing and six a higher score. Of those now hospitalized from 6 to 8 years, however, nine had a lower retest score and only three a higher score. This difference, though not significant, is contrary to Yates' (1956) hypothesis that length of institutionalization is the most important factor in inducing a decline in vocabulary performance.

Adequate assessment of the role of aging in the observed slight decline would require comparable data on normal *Ss* which, unfortunately, are unavailable. Since *all Ss* were now six years older, a measure of the effect of aging by relating age to amount of change would test for rate of change at different ages, not amount of change due to aging. Thus, although rate of change did not differ significantly with different ages in the present sample, consistent change in the entire group as a function of normal aging seems, by process of elimination, to be the most plausible explanation of the observed slight decline in verbal facility.

However subtle the change in performance on the word meaning measures was in the six years, and whatever the explanation for it, it is clear that changes even as slight as those observed in the present sample can vitiate the use of current vocabulary performance as an estimate of former ability level for use in a deterioration index. This was illustrated by the comparison of deterioration indices constructed separately from performance in 1952

and 1958, which indicated no change, although a similar index contrasting actual former ability level (1952 vocabulary) with present mental efficiency (1958 analogies and similarities) revealed a significant "deterioration" in the six-year period.

The conclusions drawn from re-examination of the use and understanding of word meanings by these chronic schizophrenics, after six additional years of hospitalization, were as follows:

1. Overall performance on word meaning measures was remarkably stable.
2. Performance on definitions tests tended to drop very slightly but consistently among the *Ss*.
3. This subtle decline in performance on definitions tests seemed more likely due to slightly decreased verbal facility or communication of knowledge, rather than to loss of knowledge per se.
4. Loss in verbal facility seemed more plausibly attributed (by process of elimination) to normal aging processes than to increased schizophrenic pathology or to length of time institutionalized.
5. A deterioration index, using the discrepancy between current vocabulary test performance and current performance on problem solving tasks, completely obscured an observed significant deterioration in performance.
6. Impairment of vocabulary performance in schizophrenics seems to be manifest as a decline coincident with and a function of the psychotic episode itself, rather than as a gradual erosion with continued illness and institutionalization.

PROGNOSTIC UTILITY OF WORD MEANING MEASURES

Word meaning measures of the type described in the previous section, though frequently used in various ways for psychodiagnosis, rarely have been tested empirically for prognostic value. In the present study it was possible to locate 47 of the 50 surviving *Ss*. Thus, an opportunity was afforded to compare directly the current adjustment of the *Ss* with their performance on the word meaning measures six years earlier.

Two criteria of current adjustment were available. First, 12 *Ss* had been discharged and were living in the community, as against 35 who were still hospitalized. Although discharge from an NP hospital may be conditioned by a great many sociological factors having little to do with the patient's psychiatric condition, it was presumed that discharge could be used as a moderately reliable criterion of significant improvement. Second, the social adequacy of each *S* in his current environment was evaluated by an experienced social worker. The Pinchak-Rollins Social Adequacy Rating Scale (SARS) was used for this purpose.

Social Adequacy Ratings[10]

The SARS provides a five-point scale for each of eight areas of adjustment. Separate forms are provided for rating adjustment in the hospital and in the community, with slight changes in wording to make the rating appropriate to the setting. Detailed instructions are given, with a summary of approximately 100 words describing the nature and scope of the variable to be rated. Each of the five points on a variable carries a sentence describing the range and types of behavior to be rated for that particular area of adjustment. Variables also are scored 1, 2, or 3 for confidence in the rating.

The eight areas of rated adequacy are given below. The wording in parentheses was substituted for that in italics when the scale was used to rate hospitalized *Ss*.

A. Responsibility for money (or canteen coupons). The extent to which the individual is realistically responsible about his money (or coupons), its source, and its use.

B. Personal appearance. The extent to which the individual has been able to maintain his personal appearance, e.g., dressing, shaving, oral hygiene, haircuts, grooming, etc., in conformity with *those of others in his environment* (expectations for other persons in society).

C. Personal habits. The extent to which the individual has been able to maintain *standards of his immediate environment* as to eating, sleeping, and bathing (social standards).

[10] We are in the debt of Louis E. Pinchak and Glen W. Rollins, Social Work Service of the Waco Veterans Administration Hospital, for use of the SARS and for the great contribution of Pinchak in locating *Ss* throughout the United States and personally supervising the SARS ratings. A comprehensive report on the SARS is in press (Pinchak & Rollins, 1959). The separate scales for adjustment in the community and in the hospital, as used in the present study, have been deposited with the ADI. Order Document No. 6318 from the ADI Auxiliary Publication Project, Photoduplication Service, Library of Congress, Washington 25, D.C., remitting in advance $1.75 for microfilm or $2.50 for photocopies. Make checks payable to Chief, Photoduplication Service, Library of Congress.

D. Responsibility for vocational (educational or industrial assignments). The extent to which the individual has been able to maintain continuing responsibility for productive work (on his assignment).

E. Social group attendance. The extent to which the individual assumes responsibility for attending (ward or hospital) social, recreational, religious, and avocational activities.

F. Social group participation. The extent to which the individual is responsible for emotional interaction in a social group (on the ward or in special hospital groups).

G. Responsibility for family and immediate associates (other patients and personnel). The extent to which the individual shows responsibility for the rights and well-being of *members of family and immediate associates in his environment at home, at church, on the job, etc.*, (other patients, ward personnel, and other associates).

H. Interpersonal relations. The extent to which the individual maintains sustained relationships with others (patients and personnel) on a person-to-person basis, demonstrating personal-emotional involvement.

The SARC was used by social workers in the Veterans Administration to rate the 35 patients in the various hospitals in which they resided. The 12 former patients were rated in their local community by social workers in the Veterans Administration and in collaborating agencies. The majority of *S*s were in Texas, but a sizable minority were scattered throughout the United States. For 12 judges independently rating the same 10 *S*s, Pinchak and Rollins (1959) report an average intercorrelation of .89 on SARS total score.

Prediction of Social Adequacy in 1958

The extremely poor prognosis for almost all of the *S*s in the present sample was one of the most immediate and striking findings in this follow-up. Only 12 of the original sample had been discharged within the six year period, and of these only one was judged to be making an adequate adjustment and one a borderline adjustment. The other 10 discharged *S*s were found to require active supervision by family and associates in order to remain in the community.

Since virtually none of the *S*s had recovered, the prognostic efficiency of the word meaning measures was tested against an extremely restricted criterion, the degree of illness within a pathological group rather than pathology vs. recovery. Despite this restriction, it may be seen in Table 4 that performance in 1952 on all of the word meaning measures showed a slight positive correlation with social adequacy in 1958, whether judged by discharge or by the SARS ratings. Although virtually none of the individual correlations were statistically significant, the fact that all 27 correlations were in the predicted positive direction strongly suggested the existence of a very slight but definite relationship.

TABLE 4

PREDICTION OF SOCIAL ADEQUACY IN 1958 FROM PERFORMANCE ON WORD MEANING MEASURES IN 1952

Word Meaning Measure	Correlation with 1952 Test Performance		
	Discharge[a] from Hospital (n = 47)	Hospital Adjustment[b] (n = 35)	Outside Adjustment[b] (n = 12)
Wechsler-Bellevue Vocabulary subtest, Form I	.31	.16	.30
Other Word Meaning Measures:			
Abstract definition, total	.35	.40*	.11
Synonym (recall) percentage correct	.26	.28	.23
Synonym (recognition) percentage correct	.27	.17	.29
Sentence construction, total correct	.17	.09	.40
Adequate similarities, total	.37	.16	.19
Abstract similarities, total	.38	.12	.35
Analogies, total correct	.32	.22	.34
Word association, relatedness[c]	.32	.19	.19

[a] Biserial r, σr_b = .198.

[b] Rank order correlation coefficients.

[c] n = 46, 34, and 12, respectively; one hospitalized *S* did not complete test.

* $p < .05$.

The deterioration index constructed from performance in 1952 did not correlate significantly with discharge from the hospital or with rated social adequacy. Age and years institutionalized correlated −.03 and .04, respectively, with rated social adequacy (n = 47).

It may be concluded that performance on the word meaning measures in 1952 showed a very slight but apparently genuine relationship to discharge from the hospital and to rated social adequacy in 1958, but that a deterioration index, using a ratio of the same measures, was unrelated to either criterion.

SUMMARY

A sample of 55 chronic paranoid schizophrenics that had been examined in 1952 with an extended battery of word meaning measures was followed up in 1958. Thirty of the 35 *Ss* still hospitalized were readministered the same test battery. Overall performance was remarkably stable over the six year period, with no significant mean change on any measure. However, performance on definitions showed a slight decline in most *Ss*. This decline seemed due to loss of verbal facility rather than to loss of knowledge, or memory failure. Also, the decline seemed associated more plausibly with normal aging than with increased schizophrenic pathology or length of time institutionalized. A deterioration index, using a ratio of WB vocabulary to analogies plus similarities was demonstrated to obscure an observed "deterioration" in performance.

The 35 hospitalized *Ss* and 12 discharged *Ss* also were rated for social adequacy in 1958. Only two of the discharged *Ss* were judged to be making an adequate adjustment. Test performance in 1952 showed an extremely low but positive relationship to discharge from the hospital and to rated social adequacy. A deterioration index, constructed from the same tests, showed no relation to either criterion.

REFERENCES

FEIFEL, H. Qualitative differences in the vocabulary responses of normals and abnormals. *Genet. psychol. Monogr.*, 1949, **39**, 151–204.

MORAN, L. J., MORAN, F. A., & BLAKE, R. R. An investigation of the vocabulary performance of schizophrenics: II. Conceptual level of definitions. *J. genet. Psychol.*, 1952, **80**, 107–132.

MORAN, L. J. Vocabulary knowledge and usage among normal and schizophrenic subjects. *Psychol. Monogr.*, 1953, **67**(20, Whole No. 370).

PINCHAK, L. E., & ROLLINS, G. W. The social adequacy rating scale, a preliminary report. In press.

YATES, A. J. The use of vocabulary in the measurement of intellectual deterioration: A review. *J. ment. Sci.*, 1956, **102**, 409–440.

(Received August 3, 1959)

Journal of Abnormal and Social Psychology
1960, Vol. 61, No. 2, 255-262

A MICROGENETIC APPROACH TO THE AMNESIC SYNDROME[1]

GEORGE A. TALLAND
Harvard Medical School

In KORSAKOFF'S psychosis the memory defect is so outstanding, so much more severe than other symptoms that the disease is customarily referred to as the amnesic syndrome. Experimental studies (Talland, 1958b, 1959a) have thrown more light on the cognitive impairment characteristic of this syndrome, suggesting dysfunctions in basic processes other than that which is broadly known as memory. Korsakoff patients function defectively in at least three major areas of cognitive processes: (*a*) in the selective operations which determine perceptual discrimination and the categorization of concepts, (*b*) in reorienting their initial attitude towards a task or situation confronting them, and (*c*) in their capacity to integrate sequential experiences into unitary structures. Quite possibly, these three phenotypically distinct processes are functionally related to one another and also to the processes by which experiences become available for voluntary recall. The tendency to isolate fragmentary events and observations would seem to be particularly relevant to the twofold amnesic disturbance of Korsakoff patients, i.e., their virtual inability to learn or recall information received and events witnessed after the onset of the disease, and their diminished capacity to remember those which had happened and which they had known before. Impairment in temporal integration was a defect characteristic of the two extreme cases of amnesia reported in the literature (Conrad, 1953; Störring, 1931), and has been experimentally demonstrated with a group of less severely damaged Korsakoff patients (Talland, 1958a). The relevance of this defect to the amnesic symptoms was suggested also by Lidz (1942) and is at least implicit in Bürger-Prinz and Kaila's (1930) graphic definition of the Korsakoff patient as consisting of cross sections without a longitudinal continuity.

In order to be available for recall, experiences must become embedded in some cognitive systems, and quite obviously this happens as a neurophysiological process in the brain. Common experience, as well as studies of such damaged functioning as occurs in the Korsakoff syndrome, shows that clear perception is not a sufficient condition for accurate recall or for voluntary recall of even approximate exactitude. Remembering depends on an additional process of embedding which is developmentally subsequent to perception. In the intact brain, the complete process of perception is likely to include the phase of fixation, but in abnormal conditions these two processes may become divorced, and perception itself may not reach full development. An incompletely structured percept is also inadequately structured within the cognitive system which subserves recall.

Perception is usually thought of as an instantaneous event, for the process rarely admits analysis into stages. The Gestalt school was particularly emphatic on the immediacy of the experience, although Bartlett (1916), Michotte (1927), and Sander (1927) demonstrated that objective analysis need not imply a return to Wundtian introspection and argued in favor of studying perception as a developmental process. Hebb's (1949) associationist theory also treats perception as a progressive process and is particularly relevant to the case of Korsakoff's syndrome, in that its neurological model offers an explanation why an incompletely structured percept should be lost to memory. Most of the experimental work and theoretical formulations directed to the developmental study of cognitive processes have been done in Sander's wake, and are now best known by Werner's (1956) translation of *Aktualgenese* as the microgenetic approach (Flavell & Draguns, 1957). The basic tenet of this approach is that perception, like concept formation or problem solving, is a process of development from the diffuse to the

[1] This research was supported in part by a Research Grant, M-767(c) from the National Institute of Mental Health, United States Public Health Service.

G. Talland is connected with the Psychology Laboratory, Department of Psychiatry, Massachusetts General Hospital, supported in part by Air Force Contract AF-49-(638)-98; the Merrill Trust; the Harrington Trust, Harvard Medical School.

structured, progressing over a microscopic temporal sequence. Normally, the successive phases of this development are indistinguishable; only the end product can be inspected. However, under suitable experimental conditions the progress of microgenesis can be brought under observation.

The purpose of the present study was to test certain hypotheses which follow from the foregoing formulation of the cognitive disorder in Korsakoff's syndrome, by investigating the microgenetic process of perception. Several findings concerning the rigid adherence of the Korsakoff patients to a particular orientation or set (Bürger-Prinz & Kaila, 1930; Grünthal, 1924; Müller-Suur, 1949; Talland & Miller, 1959) point to an impairment of their microgenetic processes, for these must be corrective as well as cumulative. The incapacity to integrate sequences of input may also extend to the order of intervals over which percepts develop microgenetically, intervals so short that memory effects are excluded from their range by definition. More particularly, the argument for an impairment of the microgenetic process of perception in the Korsakoff syndrome would be supported if the following hypotheses were to be confirmed:

1. In Korsakoff patients, under conditions of progressively increased exposures, the perception of ambiguous figures progresses towards an accurate and structured percept at a slower rate and less consistently than in controls.

2. With Korsakoff patients, successive exposure of the same ambiguous figure does not result in a more consistent or more accurate approach to perceptual structuring than does presentation repeated after intervals.

Structuring, in the context of these two hypotheses, means the integration and elaboration of parts or details into whole figures; consistent stands for the gradual development of this process, accurate for its veridicality.

METHOD

Design. The study consisted of three experiments. The purpose of the first two was to test the two hypotheses stated above, by providing the conditions for a gradual structuring of percepts: one over a more extended period of time with intervening activity, the other in a continuous process. Experiment III, in which no control group participated, attempted to sort out the effects of repeated exposure and length of exposure on the microgenetic progress of perception.

Subjects. The experimental sample consisted of 16 patients, 4 men and 12 women, diagnosed by neurologists as having Korsakoff's psychosis. They were in the chronic phase of the disease, having been at the Boston State Hospital for periods ranging from 3 to 15 years or longer. They all had a history of alcoholism prior to hospitalization which, in the first place, was for the neurological symptoms of Wernicke's encephalopathy.

There were two control groups, one consisting of 15 and the other of 10 psychiatric and neurological patients whose sociocultural background was similar to that of the Korsakoff patients, but who were otherwise not matched with them.

Material. The five composite figures shown in Fig. 1 were used in all the experiments. The first four are legendary beasts: (A) the winged bull of ancient Assyria and of the Evangelist St. Luke; (B) the hippogriff, a composite of horse, eagle, and lion; (C) merman and mermaid; (D) the Etruscan hippocamp, half horse and half dolphin. They were copied with some modifications from a book on fabulous monsters (Lum 1951). Fig. 1E is a caricature[1] of the Chevalier d'Eon, French diplomat and famed duelist, whose transvestite habits gained him considerable notoriety in London society of the 1770s (White, 1952). These inherently ambiguous designs were chosen for their relative incongruity and unfamiliarity, thus creating conditions for a more extended process of recognition by demanding the repeated checking of observations and the reconciliation of incoherent details. Photographs of the drawings, none of which was over 7 in. in its longer, nor less than 4½ in. in the shorter dimension, were mounted on cardboard and viewed by *S* binocularly in a Gerbrands tachistoscope at a distance of about 24 in. from his eyes, in the center of a field of vision of 7¾ × 7¾ in.

Procedure. In all the experiments, *S* was instructed to be alert, watch, and report what the picture was he had seen flashing in the instrument. He was told that the time for inspection would be very brief, but received no clue in advance of the type of figure he would be shown, nor subsequently any suggestion of what it may have represented. However, if he reported nothing or if his response was very meager, *E* prompted him with such nondirective questions as "Is that all?", "Anything else?" or by repeating his words with an inflection of incompleteness.

In Experiment I the composite figures were shown under the conditions of Repeated Presentation. The five figures were presented in standard sequence from A to E, first with an exposure of .1 sec., then again in the same order with exposures of .25, .5, and 1.0 sec. All 16 Korsakoff patients and 15 control *S*s participated in this procedure. Experiment II was conducted under the conditions of Successive Presentation, in which each figure was shown four times with exposures of .1, .25, .5, and 1.0 sec., before presenting the next one. Fourteen Korsakoff patients and a control group of 10 *S*s, different from those in the first control group, participated. The reason for

[1] The composite figure of the Chevalier d'Eon originally appeared in the September 1777 issue of the *London Magazine.*

A

B

C

D

E

FIG. 1. Composite figures.

A. Assyrian winged bull
B. Hippogriff
C. Merman and mermaid
D. Etruscan hippocamp
E. Chevalier d'Eon

introducing the two experimental conditions in that order was to protect against a spurious support of the second hypothesis. Since the two experiments were conducted six months apart, it was unlikely that any memory effect would carry over from the first to the second occasion in the amnesic patients. Nevertheless, if such an effect should prevail, it would not favor perceptual structuring under conditions of Repeated Presentation. In the control sample, of course, the operation of such an effect was quite probable, and this was the reason for selecting different *S*s for the two tests.

Experiment III was administered to the 14 Korsakoff patients four weeks later, using again the method of Successive Presentation. This time, each of the seven figures was shown with a different schedule; one seven times with exposures of .01, .01, .01, .1, .25, .5, and 1.0 sec.; another six times with exposures of .1, .1, .1, .25, .5, and 1.0 sec.; a third five times with exposures of .25, .25, .25, .5, and 1.0 sec.; a fourth with exposures of .5, .5, .5, and 1.0 sec.; the last three times, each for a full second. The schedule of presentation was shifted across the five figures so that in the group of 14 *S*s each design was shown three times with four schedules and twice with a fifth. Last, the figures were presented one by one outside the tachistoscope, so that *S* could inspect them for as long as he cared to before describing each. Full inspection was allowed to the control group at the end of Experiment II.

Scoring. The data used in the analysis of results were not directly derived from the experimental procedure. Responses to the figures were recorded verbatim, and subsequently tabulated under several headings. In principle, this procedure is subject to the risk of systematic error; however, the main coding categories used in this instance represented three so clearly defined universes that, after the establishment of certain guiding rules, hardly any occasion for error could occur. For responses to the first presentation of a figure the protocols were subdivided into correct, partly correct, or incorrect. Horse, for Fig. 1A was incorrect, for Fig. 1B partly correct, as was lion, eagle, or bird, but not rooster. Responses to subsequent presentations were distinguished according to their being Elaborations, Changes, or Replications of a previous description, and then divided into correct and incorrect. Elaboration included all modifications of a previous response in the direction of a structured percept, a new detail, or the correction of a detail, as well as a more specific definition instead of a general one (e. g., horse for animal, two fishes for fish). Change meant the replacement of a specific response by another which could be correct (e.g., mermaids for butterflies) or incorrect (e.g., Napoleon for George Washington). Replication could either be a verbatim repetition of a previous response, or words amounting to much the same (e.g., circus horse for trick horse), or simply the phrase "same" beyond which *S* would not expand. These three classes do not exhaust all the responses; the residual includes all instances of impoverishment, i.e., a response sparser in content yet no more specific than a previous description of the same figure, meaningless answers, comments like "black and white," and finally failure to report anything at all.

RESULTS

In the tabulation of the results only four headings were used because the instances of Change were too few to allow for further splitting; and the significance of Same is not contingent on the accuracy of the previous response, for no matter how correct that may have been it invariably allowed for considerable further Elaboration. Since the analysis is concerned with trends in the direction of perceptual structuring, the tables do not include scores based on responses to the first presentation of a figure. For the purposes of tabulation, responses to the subsequent presentations of the five composite figures are combined into single scores, and these are expressed as a percentage of the highest possible total for any one *S*, i.e., 15 in Experiments I and II.

With an exposure of .1 sec. one Korsakoff patient failed to see anything articulate in four composite figures on both first and second testing, another failed likewise with three figures on the first occasion and did not take part in the experiment with Successive presentation. Apart from these instances, defective vision or the shortness of the exposure did not interfere with the perception of some distinct figure in either experimental or control groups. Most members of the latter gave a correct or partly correct answer after the briefest exposure of Fig. 1B, C, D, and E, but only three recognized Fig. 1A as a bull or a nonspecific animal in the first and second experiment alike. The same number of Korsakoff patients responded more or less correctly to the first exposure of Fig. 1A in both Experiments I and II, as compared with half of the group on the first presentation of Fig. 1C, and about two-thirds with Fig. 1B, D, and E. Almost all of them gave a correct or partly correct description of the last four figures on subsequent presentations.

The tendency in the two groups to Elaborate, Change, or Replicate previous descriptions can be traced from the data in Table 1, and then compared for the conditions of presentation. Differences between the mean scores of the Korsakoff and control groups, between the performances of the two control groups, and of the Korsakoff patients under the two conditions of presentation have been tested for

TABLE 1

MEAN ELABORATION, CORRECT ELABORATION, CHANGE, AND REPLICATION SCORES OF THE KORSAKOFF AND CONTROL GROUPS IN EXPERIMENTS I AND II AFTER SECOND, THIRD, AND FOURTH PRESENTATIONS OF THE COMBINED FIVE COMPOSITE FIGURES

(Scores express the actual response as a percentage of the possible total)

	Elaboration			Correct Elaboration			Change			Replication		
	Kors.	Con.	t Ratio	Kors.	Con.	t Ratio	Kors.	Con.	t Ratio	Kors.	Con.	t Ratio
Repeated Presentation	41	62	2.83**	23	41	3.37**	23	11	1.16	24	18	1.10
Successive Presentation	24	59	$z = 3.48$**	17	42	$z = 2.56$**	6	13	$z = 1.60$*	65	20	$z = 3.62$**
t ratio	4.85**	.26		2.35*	.34		4.11**	.35		11.50**	.28	

* $p < .05$.
** $p < .01$.

significance, using t ratio for correlated means in the last instance, and Mosteller and Bush's (1954) version of the Mann-Whitney test when variances are not homogeneous. This test yields a standard score (z).

It is apparent that in both experimental conditions the rate of total Elaboration and of correct Elaboration is lower in the Korsakoff than in the control group, but that the former Change their description significantly less often and Replicate it more frequently only under conditions of Successive Presentation. The two control groups differ in none of these aspects of their performance, though one was presented with the composite figures *repeatedly* and the other *successively*. The Korsakoff group's total and correct Elaboration and Change were significantly higher, its Replication was significantly lower under conditions of repeated presentation.

The results of Experiment III are presented in Table 2. Neither Elaboration nor Replication seems to be a direct function of the number of successive presentations, and neither varies with the mean length of exposure in a direction which would suggest that—at any rate, within the range of 1 sec.—the briefness of exposure time accounted for the low rate at which Korsakoff patients evolved their percepts. Table 2 does not show but the records furnish further evidence in support of this negative conclusion. In the longer sequences, Elaboration occurred as often after a short as it did after a long exposure, and many of the Replications came towards the end of the series, i.e., following a longer exposure. Also, in accordance with this conclusion is the absence of significant differences between any of the mean scores in Experiment III and the corresponding scores in Experiment II.

TABLE 2

MEAN ELABORATION, CORRECT ELABORATION, CHANGE, AND REPLICATION SCORES OF THE KORSAKOFF GROUP IN EXPERIMENT III, FOLLOWING THE SECOND AND SUBSEQUENT PRESENTATIONS OF THE FIVE COMPOSITE FIGURES

(Scores express the actual response as a percentage of the possible total)

Number of successive exposures	3	4	5	6	7	All
Elaboration	18	31	21	33	31	28
Correct Elaboration	11	21	14	26	23	20
Change	7	7	5	9	4	6
Replication	68	55	64	53	56	58

Inspection outside the tachistoscope elicited from the Korsakoff patients barely more descriptive material, as judged by number of words or units of report, than tachistoscopic presentation; nevertheless, 80% of it involved some Elaboration and 76% was veridical, all the errors occurring with Fig. 1A. In the control group, 78% of the responses to prolonged inspection constituted some Elaboration of previous descriptions, all correct.

As a matter of interest it may be noted that, excepting correct Elaboration of Fig. 1A, the contribution of the five figures to the four sets of scores was remarkably even in all experiments and with both groups of *Ss*. Figures 1B and C provided the best material

for Elaboration, Fig. 1D and E for Replication, Fig. 1A for incorrect Elaboration or Change. Figure 1E also inspired many Changes, more particularly from one sex to the other, but also across the periods of history. In repeated presentations, one *S* recognized it first as Napoleon and then as George Washington, another in a different group as a gladiator and Don Juan. Other specific identifications included Queen Elizabeth, Abraham Lincoln, Martha Washington, Scots dancer with bagpipes, Chinaman, Jew with locks of hair, Spanish dancer, and knight in armor. Seven control and three Korsakoff *S*s recognized the combination of man and woman in this figure after Successive tachistoscopic presentation, and so did two additional members of either group following longer inspection. The remaining *S*s assimilated the incongruous details of the image perceived, or avoided the identification of its sex. When asked whether it was a man or woman, their answers sounded definite enough, though in some instances supported by reasoning of this type: "it must be a woman because of the feathers."

DISCUSSION

The experiments confirm the hypotheses tested. The rate at which Korsakoff patients structure ambiguous figures into percepts is slower, and under Successive Presentation it is also more often interrupted by phases of stagnation, than in the control group. Experimental conditions designed to display the microgenetic development of perception actually show less progress in Korsakoff patients than do conditions which interrupt this process; there is no difference between the effects of the two conditions on the control groups. The Korsakoff patients' failure in the progressive structuring of percepts and in the gradual correct identification of the composite figures may therefore not be attributed to a primary memory defect, for this would not be manifested in the brief intervals between two Successive Presentations of a figure, and would certainly be more effective over the longer period which elapses between Repeated Presentations, quite apart from the interference the intervening tasks might or might not produce (Talland, 1959b).

The differential results of the experiments with composite figures seem paradoxical only if the amnesic defect is regarded as the primary psychological symptom of Korsakoff's psychosis. They agree with other observations which demonstrate the inability of those patients to reorient themselves to a situation, to modify their initial perception of or attitude to it. Likewise, on account of their rigid adherence to the first percept, successive presentations, even well above the threshold of clear discrimination, make for very slight increments in structuring or, as the results of correct Elaboration show, in approximating a veridical percept. In contrast, Repeated Presentation by interrupting the first set adopted towards a composite figure, advances the process of structuring and correction at a faster and steadier rate. Of course, the interruption of a set can only indirectly advance the process of perception. In the absence of any transfer of information from one exposure of a figure to its Repeated Presentation, the *S* would only be in the same position he initially held.

It must therefore be inferred that some memory effects did operate in Korsakoff patients, making for a gradual advance in their perception both towards completeness and accuracy. The alternative explanation would attribute this progress to the increased exposure time with each Repeated Presentation. Experiment III, though not decisive on this issue, would argue in favor of the first explanation. It shows that the length of exposure time—within the range deemed suitable for exploring the microgenesis of perception—is not directly related to the rate of Elaboration or the likelihood of Replication. The responses given by the Korsakoff patients after prolonged inspection of the figures were significant not so much for their brevity or rapid closure, as for containing some elaboration in spite of the paucity of content, in almost every instance. The quality of their reports following tachistoscopic presentation could therefore not be attributed to any impediment in verbal fluency.

A genuine deficit in the microgenetic development of perception thus seems fairly well established in the amnesic syndrome. This observation would obviously gain in sig-

nificance if it were theoretically related to other, clinically more apparent, symptoms of the disease. Kragh (1955), discussing the microgenesis of perception, suggested that possibly every sequence of development must be repeated a number of times before it can acquire sufficient strength to result in full-fledged perception. He illustrated the completion of a perceptual process with a negative instance drawn from Schumann's (1907) experiment in which, after brief exposure, *S*s were known to report either that they could not recognize the word even though they had seen the letters clearly, or that they had recognized it but could not recall it. The microgenetic paradigm of perception fits neatly into Hebb's (1949) model of simple cell assemblies aggregating into phase sequences over a period of time. In this model, learning proceeds and memories become available the same way as perception occurs, i.e., by means of the establishment of facilitations in the nervous system, though at different levels of organization. The effect of these facilitatory channels is to increase the probability that certain events in the brain follow other events, that a percept attain completion, that a concept be formed, that a memory recur when sought. If it is assumed that in Korsakoff patients these facilitations fail to become established, the probabilistic model can account for those instances in which their recall, recognition, or other cognitive processes function efficiently, as well as for the many more occasions in which they do not. The model also accommodates the various observations concerning the Korsakoff patients' relative incapacity to shift expectancies, and certainly agrees with the paradigm of isolated experiences proposed above. It has the virtue of comprehensiveness, but is inherently no more convincing than an inhibitory model. Several observations of the Korsakoff patients' behavior, more particularly of the kind of verbal performance which is often referred to as confabulation, fits more closely a model of defective inhibitory mechanisms. There is no necessity to choose either model at the expense of the other, for it is reasonable to believe that a disease which interferes with the efficient functioning of facilitatory systems also damages the inhibitory processes. Neither is the microgenetic approach to cognition tied to any one neurological model; its purpose in this study was to demonstrate the applicability to the psychopathology of Korsakoff's syndrome of such purely psychological constructs as are the isolation of single experiences and the relative fixity of originally adopted attitudes.

SUMMARY

The microgenetic process of perception was examined in a group of Korsakoff patients by means of serial tachistoscopic presentation of ambiguous drawing with successively longer exposure. The drawings consisted of unfamiliar composite figures.

Under the experimental condition of successive presentation, which is the most favorable for observing this process, the gradual structuring and elaboration of percepts proceeded at a slower rate than it did when other tasks intervened between the two exposures of the same figure. Under both conditions the process of structuring and elaboration was slower and less complete with the Korsakoff patients than in control groups of psychiatric and neurological patients. The number of repeated exposures and length of inspection did not determine the rate of structuring and elaboration.

The differential effect of the Korsakoff patients' performance under the two conditions of presentation supports the view that the perceptual defect in Korsakoff's syndrome is not secondary to a memory disorder, and that an isolation of rudimentary impressions and an impaired capacity to revise attitudes are general psychological symptoms of this disease.

The findings based on this microgenetic approach and of other cognitive studies of Korsakoff's syndrome were considered in the light of Hebb's neurophysiological model.

REFERENCES

BARTLETT, F. C. An experimental study of some problems of perceiving and imaging. *Brit. J. Psychol.*, 1916, **8**, 222–266.

BÜRGER-PRINZ, H., & KAILA, M. Über die Stuktur des amnestischen Symptomenkomplexes. *Z. Neurol. Psychiat.*, 1930, **134**, 553–595.

CONRAD, K. Über einen Fall von "Minuten Gedächtins." *Arch. Psychiat.*, 1953, **190**, 471–502.

FLAVELL, J. H., & DRAGUNS, J. A microgenetic approach to perception and thought. *Psychol. Bull.*, 1957, **54**, 197–217.

GRÜNTHAL, E. Über das Symptom der Einstellungsstörung bei exogenen Psychosen. *Z. Neurol. Psychiat.*, 1924, **92**, 255–266.

HEBB, D. O. *The organization of behavior*. New York: Wiley, 1949.

KRAGH, U. *The actual-genetic model of perception-personality*. Lund: Gleerup, 1955.

LIDZ, T. The amnestic syndrome. *Arch. Neurol. Psychiat.*, 1942, **47**, 588–605.

LUM, P. *Fabulous beasts*. New York: Pantheon, 1951.

MICHOTTE, A. Rapport sur la perception des formes. *Proc. 8 Int. Congr. Psychol.*, 166–174. Groningen: Noordhoff, 1927.

MOSTELLER, R., & BUSH, R. R. Selected quantitative techniques. In G. Lindzey (Ed.), *Handbook of social psychology*. Vol. I. Cambridge, Mass: Addison-Wesley, 1954. Pp. 289–334.

MÜLLER-SUUR, H. Beitrag zur Frage des Korsakov-Syndroms und der Analyse der amnestisch-strukturellen Demenz. *Arch. Psychiat. Nerv.*, 1949, **181**, 683–711.

SANDER, F. Über Gestaltqualitäten. *Proc. 8 Int Congr. Psychol.*, 183–189. Groningen: Noordhoff, 1927.

SCHUMANN, F. Psychologie des Lesens. *Bericht über den II Kongress für experimentelle Psychologie in Würzburg*. Leipzig: Barth, 1907. Pp. 153–183.

STÖRRING, G. Über den ersten reinen Fall eines Menschen mit völligem isoliertem Verlust der Merkfähigkeit. *Arch. Psychol.*, 1931, **81**, 257–284.

TALLAND, G. A. Perception of apparent movement in in Korsakoff's psychosis. *J. Pers.*, 1958, **26**, 337–348. (a)

TALLAND, G. A. Psychological studies of Korsakoff's psychosis: II. Perceptual functions. *J. nerv. ment. Dis.*, 1958, **127**, 197–219. (b)

TALLAND, G. A. Psychological studies of Korsakoff's psychosis: III. Concept formation. *J. nerv. ment. Dis.*, 1959, **128**, 214–225. (a)

TALLAND, G. A. The interference theory of forgetting and the amnesic syndrome. *J. abnorm. soc. Psychol.*, 1959, **59**, 10–16. (b)

TALLAND, G. A., & MILLER, ALICEMARIE. Perceptual sets in Korsakoff's psychosis. *J. abnorm. soc. Psychol.*, 1959, **58**, 234–240.

WERNER, H. Microgenesis and aphasia. *J. abnorm. soc. Psychol.*, 1956, **52**, 347–353.

WHITE, T. H. *The scandalmonger*. New York: Putnam, 1952.

(Received August 6, 1959)

Journal of Abnormal and Social Psychology
1960, Vol. 61, No. 2, 263-270

THE ADAPTIVE FUNCTIONS OF WIT AND HUMOR[1]

WALTER E. O'CONNELL[2]
University of Texas

THE PURPOSE of this study[3] was to test psychoanalytic formulations pertaining to wit and humor as personality variables related to the adjustment of the individual.

While the words "wit" and "humor" have been used almost interchangeably by theorists with a nonpsychoanalytic orientation, workers with a Freudian background have tended to perceive differences between the two concepts. These distinctions appear in the feelings expressed indirectly through the jest and in the degree of personal adjustment reflected by preferences among jests.

Appreciation of humor, in the restrictive meaning offered by Freud, is associated with empathy that still is not overwhelmed by the misfortunes of others, with suitably flexible emotionality, with little use of repression, and with tolerance for oneself and others (Flugel, 1947; Freud, 1905, 1928; Grotjahn, 1957; Kramer, 1958). The form of jest labelled humor presents an objectively stressful situation in which effective flight and fight are both relatively impossible. The chief character in the Freudian humor situation then reacts to this inescapable stress with neither resignation nor overt hostility but with a humorous retort.

Wit has been regarded as a means of indirect expression for latent hostile urges. Freud was never certain whether this mechanism was essentially pathological or adaptive. Early in his thinking he correlated the appreciation of hostile wit with psychoneurotic symptoms (1905), but in later years (1921) he included it among normal adaptive mechanisms.

Harmless or nonsense wit employs the same verbal mechanisms (e.g., play on words, displacement of accent) as humor and hostile wit, but the situational settings of the other two forms are missing. Therefore in nonsense wit there is neither jesting response to adversity nor disguised expression of hostility.

Little research has been directed toward testing the Freudian concepts of wit and humor. This dearth of studies is anomalous in the light of the increased attention being given to humor as a requisite for the psychotherapist, the diagnostician, and, generally, the mature person. Occasionally, humor has been mentioned by psychoanalytically-oriented investigators but its connotations have been confused with those of wit, which has been somewhat more frequently studied (Redlich, Levine, & Sohler, 1950). Researchers have generally postulated an inverse relationship between repression and overt behavior but a direct relationship between repression and appreciation of wit. Thus, the person who "represses" hostility should not act in an overtly hostile manner, but should enjoy hostile wit.

There are a number of theoretical and experimental obstacles to study in this area. Operational measures of repression and concomitant behaviors fail to take into account the complex and dynamic theoretical background of these concepts. Theoretically a person could repress an impulse yet show such impulse derivatives as hostile behavior or appreciation of wit in his overt actions. An adequate plan for testing the relationship between repression and appreciation of wit requires manipulation of such variables as the strength and extent of repressions, the psychological "distance" of the derivatives from the repressed impulse, and the tension generated both by external stressors and by intrapsychic conflicts (Freud, 1915). Within this theoretical framework, wit could be considered as a tension-reducing mechanism, with appreciation of wit contingent upon "some" but not "too much" tension or stress. The

[1] Based upon a doctoral dissertation submitted to the University of Texas, 1958. The author is grateful to Ira Iscoe for his guidance in this research.

Presented at APA meetings in Cincinnati, Ohio, September 1959.

[2] Now at Veterans Administration Hospital, Waco, Texas.

[3] 92-page copies of complete study, including the test items, have been deposited with University Microfilms, Inc., 313 North First Street, Ann Arbor, Michigan. Order Document No. Mic 58-3993, remitting $2.00 for 35 mm. microfilm or $3.68 for 5½ by 8½ in. photocopies.

present study was planned as an initial test of such complex interrelationships.

Hypothesis I. In a situation where hostile impulses are aroused and direct means for expressing them are blocked, hostile wit might provide an indirect outlet for the hostile drives, thereby reducing to some degree both the underlying urge and the mental energy required to block entry of the impulse into consciousness and/or overt behavior. Since a sudden reduction of energy or tension is experienced as pleasurable according to psychoanalytic theory, the vehicle for the reduction (the joke) should also be appreciated. Therefore, the first prediction is that groups of subjects exposed to stress appreciate wit more than comparable groups tested under nonstressful conditions.

Hypothesis II. Psychoanalytic theory has been strongly oriented toward the psychology of the male. Hence, theoretical deductions as to the role of wit and humor in female development are difficult to find. Grotjahn (1957) stated that in this culture hostile wit is the wit of the male rather than of the female. If the show of hostility even in derivative forms is not allowed to the female, one might predict that nonsense or "harmless wit" would find more favor with women than with men.

Hypothesis III. From a review of the development of the Freudian concept of wit one can make certain predictions about preferences for hostile wit, depending upon the interaction of the variables of adjustment (internal stress) and presence or absence of objective stress. The maladjusted person, presumably with strong unacceptable hostile impulses, employs the facade of the hostile joke for pleasurable and partial impulse gratification. But when the external situation involves a hostile atmosphere the repressions of the maladjusted person are exaggerated to such an extent that disguised release is not possible. On the other hand, the more adjusted individual, supposedly not burdened with excessive repressions (or intrapsychic tensions) can enjoy hostile wit mechanisms for tension reduction when objective stresses increase. Thus, it is predicted that under nonstressful conditions the maladjusted person should appreciate hostile wit more than the well adjusted person, but when stressors are applied, the well adjusted subject should appreciate hostile wit more than the maladjusted person.

Hypothesis IV. In contrast to his attitude toward wit, Freud always extolled the humorist as disdaining the use of excessive repressions, and he depicted humor as the most adaptive nonpathological defense. Other writers have concurred with Freud's picture of the humorist as being mature, nonhostile, and loving toward himself and others. One would predict, then, more frequent appreciation of humor in well adjusted persons.

Hypothesis V. In comparison with humor and hostile wit, less has been written on the subject of nonsense or harmless wit. This form of the comic, resting more upon the techniques of wit than upon savings in countercathexis, was given less emphasis in Freud's theoretical system (1905). There is some consensus that a person who can allow himself relaxation and release through the appreciation of nonsense wit is not overwhelmed by his problems (Kris, 1952; Levine & Redlich, 1955). Therefore, the final hypothesis states that a well adjusted person appreciates nonsense wit more than a poorly adjusted one.

METHOD

Selection of Independent Test Variable

In order to test Hypotheses III, IV, and V, a measure was needed that would separate the *Ss* into groups according to their level of adjustment. The Self-Activity Inventory (SAI), which gives an absolute discrepancy score between the *Ss*' self-rating and his ideal rating on 54 items relating to psychological needs, was selected as the independent test variable (Worchel, 1957). The rationale for the use of the absolute discrepancy score between self and ideal (S-I score) is provided by self-theory (Rogers, 1954), according to which tension is experienced when a person is aware of the discrepancy between his self-concept and his ideal concept. Therefore, it was predicted that the greater the discrepancy (high S-I score) the greater the anxiety or tension experienced. A number of studies have indicated that the S-I score is inversely related to the level of adjustment and of stress tolerance, and significantly correlated with psychometric tests of anxiety (Worchel, 1957). The reliability of this scale by a test-retest method over an eight-week interval was .79 for the self-concept and .72 for the ideal concept.

Selection of Dependent Test Variable

The rationale for choice of a dependent variable arises from Freud's belief that passive appreciation of wit and humor echoes the psychic states of the jest

maker. Thus, a test for the appreciation of wit and humor should provide an indicator of the defensive use of these devices. Two existing tests have attracted considerable interest as research tools in the field of the comic but neither follow the Freudian paradigm of humor (Cattell & Luborsky, 1947; Redlich, Levine, & Sohler, 1950). For the present study a Wit and Humor Appreciation Test was developed around forms of jests that follow deductions from psychoanalytic theory.

Test construction required the suitable definition of wit and humor, selection of a pool of items that would reasonably be subsumed under these categories, and decision as to the type of response to be recorded.

A pool of 65 jokes were assembled and typed separately on filing cards. Eleven judges (8 male and 3 female clinical psychologists) were asked to classify the 65 items in private as Humor, Hostile Wit, or Nonsense Wit. To avoid idiosyncratic definitions of these concepts, each judge was provided with a resume of Freud's ideas concerning wit and humor. Of the original pool of 65 items, 32 were unanimously classified as belonging in one of the three categories. Since 10 were in the humor category and 11 were categorized both as Hostile Wit and as Nonsense Wit, one item was eliminated randomly from each of the last two classes. At this point, there was an assortment of 30 jokes (10 each of Humor, Nonsense Wit, and Hostile Wit) available for inclusion in the Wit and Humor Appreciation Test (WHAT). Examples of items are as follows:

> In various times punishment has been cruel and frequently ineffective in arousing fear. There were men who endured the greatest tortures without bowing to its influence. One man who was to be hanged was carried to the gallows in a basket by four guards. He was carried because his arms and legs had been cut off. In this situation he called laughingly to the crowd present: "I never thought that a poor devil like me would be carried about by four servants." (Humor)
>
> While Emperor Augustus was traveling through his domains, he noticed a man in the crowds who bore a striking resemblance to himself. He beckoned to him to come over and asked:
>
> "Was your mother ever employed in my home?"
>
> "No, sire," replied the man, "but my father was." (Hostile Wit)
>
> Description of a cow followed by a couple of ducks: Milk and Quackers. (Nonsense Wit)

Ss were requested to record their preference for each joke on a scale from 0 ("dislike very much") to 4 ("like very much"). The rationale for such a scale of preference comes from psychoanalytic studies of wit and humor that regard "implicit laughter" and appreciation as the key response, rather than more overt measures.

Several pilot studies were conducted to measure the effect of extraneous variables upon test responses. To minimize testing difficulties random presentation was arranged, i.e., six jokes on each of five sheets, the sheets in random order for each *S*; on each sheet two items from each wit and humor category, counterbalanced for position on the page.

Subjects

The 332 *Ss* were college students, essentially homogeneous in regard to age and class, enrolled in seven introductory courses in psychology, sociology, or anthropology.

Procedure

Two test scores were needed for each *S* in order to evaluate hypotheses relevant to adjustment and appreciation of wit and humor. To test the predictions concerning appreciation of wit and humor under stress, certain classes were designated as stress groups, and others as nonstress groups. For the stress groups, two classes with a total of 150 *Ss* were selected, both sections in introductory anthropology meeting 2 hours apart. The nonstress groups contained 182 *Ss* from five intact classes. The testing of both groups was within the regular classroom hour in two consecutive class sessions.

For both stress and nonstress groups the *E* presented the SAI, supposedly as part of the routine testing procedure of the University, and returned during the next class session with the Wit and Humor Appreciation Test.

For the two stress classes, a faculty member accompanied the *E* on the second day and became the instigator of both failure-stress and insult techniques. He introduced himself as a professor of clinical psychology who was shocked and appalled at the attitude of the class in taking the SAI the previous class period. The results of the test, he said, pointed out that the members were either lying about themselves or were grossly maladjusted. He also stated that since the tests reflected a marked lack of adjustment, the majority of the class should give serious consideration either to leaving college or to changing their majors. After reiterating the insults and threats a few times, the instigator informed the students that they would be given one more test, and that they should display some maturity in taking it, rather than performing like high school freshmen. They were also admonished against communicating with their neighbors in any way. The *E* distributed the WHAT immediately upon the conclusion of the instigator's remarks.

After the tests were completed and collected, the aim of the experiment was explained to the *Ss*. To aid in the release of whatever tensions might have accumulated, about 15 min. were spent encouraging the *Ss* to give their views and feelings regarding the stress experience.

In order to provide the *E* with objective evidence of class behavior to supplement his own observations, eight undergraduate students were asked to observe and note in writing the reactions of the classes. All eight observers were unaware of the purpose of the experiment but concurred unanimously in their views as to the widespread feelings of anxiety and suppressed hostility in these stressed classes.

RESULTS

Hypothesis I stated that the groups of persons exposed to stress would appreciate wit

TABLE 1
MEAN APPRECIATION RESPONSES FOR WIT AND HUMOR ITEMS

Sex	Humor			Hostile Wit		Nonsense Wit		Total	
	N	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*
Male	196	17.7	5.81	24.6	6.12	17.2	6.60	59.5	15.92
Female	136	17.9	4.68	22.7	6.81	20.2	6.75	60.8	16.43
t		0.36		2.63*		4.11**		1.18	

* $p < .01$, one-tailed test.
** $p < .001$, one-tailed test.

more than comparable groups tested under nonstress conditions. The mean scores for the stress and nonstress groups on Hostile Wit as well as Nonsense Wit and Humor were quite similar under both conditions. Hypothesis I, therefore, was not confirmed.

Hypothesis II concerned another group difference, that of a sex difference in preference for Hostile as compared with Nonsense Wit. Statistical analysis confirmed the hypothesis that males appreciate Hostile Wit more than do females ($t = 2.63$, $p < .01$, one-tailed test), while women favor Nonsense Wit ($t = 4.11$, $p < .001$).

The remaining hypotheses were formulated with reference to the S-I scores of adjustment. In order to compare the mean variables of S-I scores, sex, and the presence or absence of stress, a $2 \times 2 \times 2$ analysis of variance design was selected. To assign *Ss* to the required eight groups, all S-I scores were plotted in ascending order. From this total of 332 scores, 111 *Ss* were grouped in both high and low S-I groups, to test the extremes of adjustment in a normal group. The low group included S-I scores from 0 to 39 ($M = 25.6$, $s = 8.12$), while the high S-I groups had scores from 57 to 122 ($M = 75.1$, $s = 11.96$). From both S-I groups, stressed and nonstressed *Ss* were separated, then these four groups were split into male and female groups. The smallest number of *Ss* for any of the eight groups was 17 in the high S-I, nonstressed, female group; hence this number became the standard number of *Ss* for each group. To arrive at an *N* of 17 in the seven remaining groups, *Ss* with the highest S-I scores were eliminated from the four low S-I groups, while those with the lowest S-I discrepancies were removed from the four high S-I groups.

Hypothesis III stated that the maladjusted *Ss* (as measured by high S-I scores) would appreciate Hostile Wit more than those of good adjustment (low S-I) under nonstress, but that the relationship would be reversed under stressful conditions. For the dependent variable appreciation of Hostile Wit (Table 2), the simple interaction of S-I × Stressor ($F = 6.23$, $p < .05$) indicated that the low S-I group under stress and the high S-I group under nonstress appreciated Hostile Wit more than the combination of low S-I in nonstress and high S-I in stress situations. Hypothesis III can be considered to be confirmed by this interaction. There is also another simple interaction that was significant, S-I × Sex ($F = 9.51$, $p < .01$). This interaction showed that women with low S-I scores and men with high S-I scores preferred Hostile Wit more than the opposite combination of high S-I females and low S-I males. The triple interaction of S-I × Stress × Sex was likewise significant at the .01 level ($F = 9.33$). When tested under nonstress or "resting" conditions the high S-I males and the low S-I females both showed stronger preference for Hostile Wit than the contrasting groups. The two extremes under this situation are the very high hostile preference for the high S-I man and the very low appreciation for Hostile Wit evinced by the high S-I woman. With the introduction of stress, the low S-I men appreciated Hostile Wit more than the high S-I men. Among the women, it was the low S-I group that appreciated Hostile Wit under this condition.

TABLE 2
ANALYSIS OF VARIANCE OF PREFERENCE FOR HOSTILE WIT

Source of Variation	*df*	Mean Square	*F*
S-I Discrepancy	1	31.0	1.01
Stressor Conditions	1	3.1	
Sex	1	45.8	1.50
S-I × Stressor	1	190.5	6.23*
S-I × Sex	1	291.0	9.51**
Stressor × Sex	1	48.2	1.58
S-I × Stressor × Sex	1	285.4	9.33**
Error	128	30.6	
Total	135		

* Significant at the .05 level.
** Significant at the .01 level.

TABLE 3
ANALYSIS OF VARIANCE OF PREFERENCE FOR HUMOR

Source of Variation	df	Mean Square	F
S-I Discrepancy	1	176.7	7.21**
Stressor Conditions	1	69.2	2.82
Sex	1	0.3	
S-I × Stressor	1	6.0	
S-I × Sex	1	45.8	1.90
Stressor × Sex	1	1.5	
S-I × Stressor × Sex	1	43.5	1.81
Error	128	24.6	
Total	135		

** Significant at the .01 level.

TABLE 4
ANALYSIS OF VARIANCE OF PREFERENCE FOR NONSENSE WIT

Source of Variation	df	Mean Square	F
S-I Discrepancy	1	47.1	1.20
Stressor Conditions	1	2.3	
Sex	1	423.5	10.83**
S-I × Stressor	1	1.2	
S-I × Sex	1	95.6	2.44
Stressor × Sex	1	67.9	1.74
S-I × Stressor × Sex	1	49.4	1.26
Error	128	39.1	
Total	135		

** Significant at .01 level.

Hypothesis IV stated that well adjusted individuals have a greater appreciation for humor than do maladjusted people. One *F* ratio, that of S-I scores, was large enough to be significant. That ratio was in line with the prediction of the hypothesis, and reflects the differential in preference for Humor of the low and high S-I groups (Means of 18.8 and 16.5). Thus, Hypothesis IV can be considered confirmed.

The appreciation of Nonsense Wit was the topic of Hypothesis V, the assumption being that those high in adjustment would appreciate Nonsense Wit more than the less adjusted individuals. The resulting *F* ratio did not approach significance at the .05 level. The sex difference was the only significant *F* ratio, with means of 17.0 and 20.6 for men and women, respectively.

To test the reliability of each component of the dependent variable two classes were randomly selected, 43 males and 27 females. By the split-half method reliabilities (corrected, Spearman-Brown) for the component variables were: Nonsense Wit, .84; Humor, .77; and Hostile Wit, .61. While not extremely high, the reliabilities are acceptable when it is considered that the population was fairly homogeneous and that there were only five items in each half of the test.

DISCUSSION

The results of this study show that appreciation of wit and humor is amenable to reliable measurement, and that groups separated on the basis of sex, stress, and adjustment have recourse to these devices under differing circumstances. It, therefore, seems likely that attempts to explain wit and humor preferences by one factor alone give an incomplete picture of the dynamics of wit and humor.

The first two hypotheses were concerned with testing stress and sex differences in the appreciation of wit and humor. As for Hypothesis I, no significant differences were found in wit and humor preferences between groups under stress and nonstress conditions. Hypothesis II was confirmed since males were found to appreciate Hostile Wit significantly more than did females. On the other hand, women liked Nonsense Wit far more than did men. This finding does not differ greatly from the results of the study of a college population by Landis and Ross (1933). In that study, men appreciated "repression" jokes, while the women preferred "nonsense" jokes. Since the populations tested in both studies were quite similar, although separated by 25 years in time, the sex difference in appreciation of wit appears to be rather stable. To what degree one can generalize beyond the youth of the middle class to different ages and classes is a problem for future research. It may be the expected pattern for the college male to enjoy the release of Hostile Wit while the college female appreciates the relatively motiveless pleasure of nonsense wit.

When the personality variable of adjustment was added, more interesting interactions were noted. Thus, there is some experimental evidence that measures of adjustment may be important in predicting appreciation of wit and humor. From the interaction of the self-ideal discrepancy with the measurement of

Hostile Wit certain features of Hostile Wit become clearer. That is, Hostile Wit appears to be predominantly a male reaction, and in everyday situations it is the more maladjusted male who is disposed to employ it. Yet under stress conditions the maladjusted males did not utilize this defense as much as did the well-adjusted males. These results are congruent with psychoanalytic theory, according to which the maladjusted person represses hostility, but can make use of the verbal distortions of Hostile Wit for partial release so long as the objective situation is not too stressful (Freud, 1915).

When the external situation has hostile implications, the maladjusted person seems to intensify his repressive efforts to the degree that wit is not so ready a release. On the other hand, the adjusted person can make greater use of Hostile Wit under situations of stress, since the additional weight of insult and failure threats is not so disabling to him. The findings regarding Hostile Wit also dovetail with the views of Alper (1957), who implied that the maladjusted individual uses defenses in ordinary situations to such an extreme extent that under stress he cannot call upon more intensive defense and is, therefore, more prone to disorganization. On the other hand, the well adjusted person can employ such defensive reserve under stress, since his stability is not strained during "resting" states.

Findings with hostile wit make possible a better interpretation of previous studies of wit. Certain researchers were ready to discount Freudian theory because of a positive relation noted between appreciation of Wit and hostile behaviors (Byrne, 1956; Luborsky & Cattell, 1947). In these studies, the key variables of adjustment and stress were not measured, and it was erroneously implied that psychoanalytic theory considers that the neurotic is not hostile (e.g., inverse relationship between repression and overt behavior). In this study it was found that maladjusted people, as defined by their self-ideal discrepancy scores, enjoy hostile wit under nonstressful situations. The findings of such investigators as Barry (1928) and Murray (1934) are in agreement. Similarly, Byrne (1956) discovered that among maladjusted hospitalized male patients seemingly not under stress, openly and subtly hostile subjects liked hostile cartoons.

Two *E*s seem to have allowed stress to enter their studies, although the designs did not originally call for it. Doris and Fierman (1955) found a significant difference in preference for hostile cartoons between *S*s who showed high and those who had low anxiety, but *only* when an opposite-sexed tester was present. In such a situation, the highly anxious *S* showed significantly less appreciation for hostile cartoons. It seems possible that a highly anxious person might be sufficiently stressed by the presence of a person of the opposite sex as a tester to lose his zest for hostile cartoons. Frankel (1953) believed she learned the problem areas of *S*s with the Blacky Test; she then administered jokes tapping such areas of maladjustment. She found that those *S*s judged to be maladjusted did not appreciate jokes reflecting the discovered problem area. Here again is the addition of stressor, this time specific to the area of maladjustment, accompanied by the lowering of appreciation for hostile wit. Frankel's method differed from that of the present study in that there was no attempt in this study to find jokes that were directed at individual problem areas.

Females seem to be an exception to the predictions of present theories. In the present experiment, it was the well adjusted girls who were able to take advantage of Hostile Wit under both stress and nonstress conditions. Perhaps if our culture frowns upon hostile wit for women, only certain types of women rely upon this mechanism for tension reduction.

It was found that, as predicted, well adjusted *S*s appreciated humor more than did the poorly adjusted. Since the responses to humor were not significantly related to sex or the presence of a stressor, and since the response to wit was more affected by the interaction factors, one might make the tentative assumption that appreciation of humor is a more stable personality characteristic than appreciation of wit.

One of the major weaknesses in this experiment pertains to the use of the S-I score as a measure of adjustment. At this point in the development of phenomenal measures, the evidence for the S-I discrepancy as a measure of the quality of a person's adjustment is not

impressive. Further research is required before the question of the validity of the S-I as an indicator of adjustment is satisfactorily answered. Perhaps for the present a more guarded label for the S-I would be one of "phenomenal self-esteem." According to the Y phenomenon (Rogers, 1954) a small discrepancy between self and ideal scores could signify either a fairly successful adjustment or a highly defensive maladjustment. In the *E*'s opinion there is some doubt about the cogency of the psychological principle in this case. The low self-ideal discrepancies reported here (Mean 25.6) are not at a pathological extreme and average almost eight points higher than low self-ideal groups used elsewhere (Miller & Worchel, 1956). There is no doubt that a study of theoretical importance could be undertaken with *S*'s giving extremely congruent self-ideal ratings (e.g., discrepancy of less than 10 points). Such a group might be characterized by extreme repression, with very limited appreciation of humor (and possibly also of hostile wit).

An item analysis of the WHAT has raised further questions that make some of the initial findings inconclusive. Within the broad categories of wit and humor sex differences in preference have been discovered depending upon the amount and direction of hostility present in the jest and the object toward which the hostility is directed. These findings, to be reported later, have led the investigator to research in progress concerning the dimensions of Freudian humor.

Further studies of individual reactions under various types of stress are also required to discover whether wit and humor have abreactive value by themselves alone or only in combination with other tension-reducing behaviors.

SUMMARY

This study was undertaken to test the possible adaptive values of the Freudian mechanisms of Humor, Hostile Wit, and Nonsense (or "harmless") Wit, for *S*s grouped according to self-ideal discrepancies, sex difference, and the presence or absence of an external stressor.

Three hundred and thirty-two *S*s indicated their self- and ideal-adjustment on Worchel's Self-Activity Inventory. The *S*s then were presented with a Wit and Humor Appreciation Test, some under insult and failure (the stressor), some under nonstressful conditions.

Five hypotheses relevant to wit and humor were tested:

I. Groups exposed to the stressor appreciate wit and humor jokes more than comparable groups that are not tested under stress. This hypothesis was not confirmed since the stressed and nonstressed groups showed no significant differences in response to the test jokes.

II. Men appreciate Hostile Wit more than women do, while women prefer Nonsense Wit more than do men. This hypothesis was confirmed with resulting probabilities beyond the .01 point.

For the remaining three hypotheses a $2 \times 2 \times 2$ analysis of variance was computed, with eight groups of 17 *S*s each, differentiated on the main variables of high and low self-ideal discrepancies, sex, and stressor—no stressor.

III. Maladjusted people appreciate hostile wit more than do well adjusted under nonstressful conditions. When a stessor is applied, well adjusted people appreciate Hostile Wit more than do the maladjusted. Hypothesis III was confirmed for the men only. The women showed a different pattern with the women of low self-ideal discrepancy (well adjusted) appreciating Hostile Wit more than the women of high self-ideal discrepancy (maladjusted), under both stressor and nonstressor conditions.

IV. The well adjusted person has a greater appreciation for humor than does the poorly adjusted person. The *F* ratio for the adjustment variable was significant at the .01 level, thus the hypothesis was confirmed.

V. The well adjusted person has a greater appreciation for Nonsense Wit than does the poorly adjusted person. This hypothesis was not confirmed.

It is concluded that the Freudian distinction between wit and humor is a valid one, with wit possibly operating to reduce tension under differing circumstances depending upon the interaction of adjustment, stress, and sex factors.

REFERENCES

ALPER, THELMA G. Predicting the direction of selective recall: Its relation to ego-strength and n achievement. *J. abnorm. soc. Psychol.*, **55**, 1957, 149–165.

BARRY, W. The role of subject-matter in individual differences in humor. *J. genet. Psychol.*, **35**, 1928, 112–128.

BYRNE, D. The relationship between humor and the expression of hostility. *J. abnorm. soc. Psychol.*, **53**, 1956, 84–89.

CATTELL, R., & LUBORSKY, L. Personality factors in response to humor. *J. abnorm. soc. Psychol.*, **42**, 1947, 402–421.

DORIS, J., & FIERMAN, ELLA. Humor and anxiety. *J. abnorm. soc. Psychol.*, **53**, 1955, 59–62.

FLUGEL, J. *Man, morals, and society*. New York: International Univer. Press, 1947.

FRANKEL, ESTER. An experimental study of psychoanalytic theories of humor. Unpublished doctoral dissertation, Ann Arbor: Univer. Michigan, 1953.

FREUD, S. Humor. *Int. J. Psychoanal.*, **9**, 1928, 1–6.

FREUD, S. Wit and its relation to the unconscious. (Originally published 1905) In A. Brill (Ed.), *The basic writings of Sigmund Freud*. New York: Modern Library, 1938. Pp. 694.

FREUD, S. Group psychology. (Originally published 1921) In J. Rickman & C. Brenner, *A general selection from the works of Sigmund Freud*. New York: Doubleday, Anchor Books, 1947. Pp. 169–209.

FREUD, S. Repression. (Originally published 1915) In *Collected papers*. Vol. III. London: Hogarth Press, 1950. Pp. 84–97.

GROTJAHN, M. *Beyond laughter*. New York: McGraw-Hill, 1957.

KRAMER, P. One of the preoedipal roots of the superego. *J. Amer. Psychoanal. Ass.*, **6**, 1958, 38–46.

KRIS, E. *Psychoanalytic explorations in art*. New York: International Univer. Press, 1952.

LANDIS, C., & ROSS, J. Humor and its relation to other personality tests. *J. soc. Psychol.*, **4**, 1933, 158–175.

LEVINE, J., & REDLICH, F. Failure to understand humor. *Psychoanal. Quart.*, **24**, 1955, 560–572.

LUBORSKY, L., & CATTELL, R. The validation of personality factors in humor. *J. Pers.*, **15**, 1947, 283–291.

MILLER, K., & WORCHEL, P. The effect of need-achievement and self-ideal discrepancy on performance. *J. Pers.*, **25**, 1956, 176–190.

MURRAY, H. The psychology of humor. *J. abnorm. soc. Psychol.*, **29**, 1934, 66–81.

REDLICH, F., LEVINE, J., & SOHLER, T. A Mirth Response Test: Preliminary report on a psychodiagnostic technique utilizing dynamics of humor, *Amer. J. Orthopsychiat.*, **21**, 1951, 717–734.

ROGERS, C., & DIAMOND, ROSALIND. (Eds.) *Psychotherapy and personality change*. Chicago: Univer. Chicago Press, 1954.

WORCHEL, P. Adaptability screening of flying personnel: Development of a self-concept inventory for predicting maladjustment. *USAF Sch. Aviat. Med. Rep.*, 1957, Randolph AFB 56–62.

(Received August 6, 1959)

Journal of Abnormal and Social Psychology
1960, Vol. 61, No. 2, 271–275

THE DEVELOPMENT OF MEANING IN CHILDREN'S DESCRIPTIVE TERMS[1]

SUSAN M. ERVIN AND GARRETT FOSTER[2]
University of California, Berkeley

CHILDREN frequently confuse the names for physical dimensions. If two objects differ in size, they may say one is STRONGER than the other. If they differ in weight, one may be called BIGGER than the other. This is, of course, just what one would expect in the early stages of learning if size, weight, and strength are empirically correlated.

Osgood, Suci, and Tannenbaum (1957) have found correlations of a similar sort in examining the structure of meanings measured by semantic differential scales. Three factors have been found repeatedly—Evaluation, Potency, and Activity. Little attention has been given to the development or origin of these factors. Inspection of the scales defining each factor suggests that two conditions would create correlations. One is verbal conjunction; if whatever is said to be GOOD is also said to be FAIR the scales will be correlated through verbal associations. Secondly, "ecological covariation" (Brunswik, 1947) exists for certain sensory dimensions. We would thus expect BIG, HEAVY, and STRONG to be correlated for all cultural groups.

It is clear that adults can discriminate the sensory dimensions that children confuse verbally. On the semantic differential, however, they are normally asked to extend terms metaphorically, as in judging the size and weight of FREE PRESS or EDUCATION. Even on the semantic differential, the correlation of size and weight scales can be destroyed by inclusion of items like DIAMOND and MIST where a check with sense experience is possible and the usual trait correlation is reversed. To the extent that the semantic differential reflects covariation in experience of traits that are logically independent, we would expect that there would be an increase with age in denotative discrimination of the terms that are correlated on the differential. In the following study, age changes in children's verbal confusions are examined.

The more highly correlated two attributes are, the less probable are encounters with discrepant instances. It is useful to note three different variants on this situation. One category may constitute a subclass of a larger category. Thus, if 98% of a child's encounters with men, in which there is direct address or verbal reference to a man, involve his father, we would expect that the child might at first call all adult males DADDY. Discrepant instances would at first be too few for a differentiation to take place. If there were two adult males in the family such an extension would be unlikely. Thus Leopold (1939) noted that while his daughter called all men PAPA, women had individual names and there was no word for FRAU.

The subclass of a hierarchy of classes is actually an extreme instance of the second variant of correlation, a partial overlap of two classes. The degree of correlation or overlap should predict the probability of two terms being confused. Thus, communism and atheism might be confused by those unaware of discrepancies such as religious communist settlements. From a matrix showing the probability of being right in applying the term COMMUNIST to an atheist, we can see that the higher the correlation of the two terms, the greater the probability of being right, and the greater the likelihood that the two ideologies are called by one term. There are other features that enter into the failure to differentiate terms. One is the relative size of the two categories. If there are more atheists than communists, one is more likely to be wrong in calling an atheist a COMMUNIST than calling a communist an ATHEIST. The extreme case would be that in which the whole size of the class of communists is equal to the overlapping class. That is, for example, all communists are atheists but the reverse is not true. This case is identical with the one cited earlier of hierarchical classes.

[1] The authors acknowledge with gratitude the cooperation of Matt Griffeath, Principal, and the teachers of the Durant School in Oakland, California, in providing facilities and *S*s.

[2] Now at the University of Texas.

Other relevant factors are the perceptibility of an attribute or class, the consequences of correct and incorrect class discrimination (which may not be the same for the classes involved), the frequency of the terms in usage, and degree of logical independence.

If the dimensions of reference are continuous attributes rather than classes, then the relation may be described by a scatterplot rather than a matrix of frequencies. The same observations apply; the probability of being correct in saying that a bigger object is HEAVIER is a function of the attribute correlations.

In adult speakers of English, the differentiation of the attributes weight, size, and strength is such that if speakers can make appropriate tests, they are unlikely to say that the heavier of two like-sized objects is BIGGER. There are, however, situations in which the attribute extension in this simple physical case is appropriate. One is in the situation of prediction, where a value on one attribute only is known. Then it becomes useful to be able to predict the probable value on the other attribute. The second situation is one in which metaphor is exploited in verbal or pictorial communication, and one attribute may be used to suggest another.

Thus, we would expect that correlated attributes would appear in experienced speakers' usage, in situations of prediction and metaphorical extension, but not in denotation where the evidence for attribute discrimination is available.

In new learners, however, one term may apply to both attributes, which are not in fact discriminated, or both terms may appear as interchangeable synonyms for the two undiscriminated attributes.

In the following study two semantic differential factors, Evaluation and Potency, are presented as far as possible in conditions requiring denotative discrimination of three attribute expressions of each. A reduction with age is predicted with respect to the use of the wrong terms when a difference in only one attribute of the correlated set is present.

METHOD

Subjects. There were two groups of *S*s, a group of 16 male and 17 female first grade children, and a group of 18 male and 18 female sixth graders. Both groups were from the same school in a lower socioeconomic Negro district. To reduce variability, the extremely bright and extremely dull children were excluded, the criterion being the teacher's rating in the first grade, and deviations of 20 points from the norm on available IQ tests in the sixth grade.

Materials. In Part I of the experiment, materials were selected to vary successively three of the dimensions on Osgood's potency factor. These three dimensions—weight, strength, and size—had been found to have loadings of .62 with the rotated factor analysis, involving concepts rated against scales, and coordinates of 1.68, 1.81, and 1.76, respectively, on another analysis of scales judged against scales. These were the largest components of the factor in each case.

The objects used were: (*a*) opaque salt shakers identical but for weight, (*b*) opaque jars identical but for weight, (*c*) cork balls differing in size, (*d*) sterofoam balls differing in size, (*e*) a pair of insulated wires with the wire removed from the middle third of one, leaving it flexible, and (*f*) a dry sponge and a damp one matched in size.

In Part II drawings of a girl's face were used to represent three of four dimensions representing Osgood's evaluative factor. The pictured dimensions were CLEAN-DIRTY, HAPPY-SAD, and PRETTY-UGLY. A fourth was included in the questioning: GOOD-BAD. These had loadings, respectively, of .82, .76, .86 (BEAUTIFUL-UGLY), and .88, and coordinates of 2.38, 2.09, 2.40, 2.29 on the first factor.

Procedure. All of the *S*s were individually tested with the following questions:

> I would like to ask you some questions and you can give me the correct answers. . . OK? [For first graders:] It's kind of a game and lots of fun. First I'm going to ask about these objects. [2 objects contrasting in weight put in subject's hands.] Is one of these heavier and one lighter or are they both the same weight? [If says different] Which is heavier? Is one black and one white or are they both the same color? Is one bigger and one smaller or are they both the same size? [If says different] Which is bigger? Is one stronger and one weaker or are they both the same strength? [If says different] Which is stronger?

If the child failed to indicate the item that was heavier on the first question, he was eliminated from the rest of Part I. The second question, to which the answer was "the same" was to control set. A similar series of questions was asked about all the items, starting with the actual contrast as a screening question. One of the six possible key questions was omitted, concerning perception of weight in objects differing in size. If weight in the objects were controlled, the smaller object might be called heavier on the grounds of its scale weight. Because of the ambiguity of the term HEAVY applied to objects differing in size, the question was omitted.

In the analysis the percentages were computed over the whole set of responses, which included two for each question for each child because of the double set of materials. The *N*s used in the significance tests were for the actual number of children tested.

In Part II a similar procedure was followed, with the control question "Does one picture have red hair and one black hair or are they both the same?" The ques-

tions and pictures were rotated, with every fourth question a control question.

RESULTS

In Table 1 it can be seen that between 39 and 66% of the first-grade children offered contaminated responses for the various dimensions of the physical materials. In the sixth grade the proportion was reduced to a range between 20 and 44%. While none of the individual changes in percentage was significant, there was a reduction for every comparison, including the subgroups by sex. A sign test is significant at the .01 level. The dimension that changes the least is the response that the heavier object is STRONGER; this is also the statement of highest frequency at both ages for both sexes. Since strength is less evident than weight, it is possible that this particular inference would continue in adult *S*s. The inference that a bigger object is STRONGER is next in probability, supporting the notion that it is the inferred character of strength that is involved. Size is least often presented as a contaminated response, and it is also the most obvious.

It might be thought that the reduction in the sixth grade is due to a tendency to be more careful about differentiating at all, and thus to an increase in *same* responses. This was not the case. While the percentage of reversals was relatively low, it increased in the sixth grade. This increase occurred in stating that the heavier was SMALLER, that the stronger was SMALLER and LIGHTER. In the latter case these frequencies probably arise because a wet sponge is usually larger, and contains water. Only boys gave the last reversal, 22% of the boys in the sixth grade saying the stronger was LIGHTER, and no girls. Nine percent of the boys and 28% of the girls said the stronger was SMALLER. In the last case this was a larger proportion than those saying it was larger. It might be argued, then, that the reduction in the last two cases was due to sophistication with respect to one of the objects used.

On the study of the evaluative dimension using faces, almost no reversals occurred at either age. The age differences on Table 2 are striking. It may be noted that they do not occur markedly on three cases. Two of these refer to CLEAN. The frequency of children saying one child was CLEANER than the other for the other attributes was small at both ages. Clean may be said to be the most visible of the attributes. In fact it could be argued that it is the only one with a clear-cut physical criterion.

The third instance of lack of marked change was one in which the proportions were very high at both ages. The smiling face was said to be PRETTIER.

With the CLEANER dimension excluded, the range in the first grade is between 42 and 97% offering a contaminated response. The highest are those offering the smiling face as more GOOD and the cleaner face as PRETTIER. All of the first grade boys gave these responses. In the sixth grade the range is between 14 and 75% with the highest now being the smiling face which is seen as PRETTIER about as often as it was in the first grade. Thus, there is a shift in responses as with the physical attributes, but it is markedly different for the different attributes.

TABLE 1

PROPORTION ASCRIBING CORRELATED ATTRIBUTE DIFFERENCES TO OBJECTS

Stimulus Differences	Response	First Grade			Sixth Grade			
		Boys	Girls	Total	Boys	Girls	Total	Differences
Heavy	BIG	40.6	36.0[a]	38.6[a]	16.7	25.0	20.8	17.8
Strong	BIG	31.2	52.0	39.7	30.6	25.0	27.8	11.9
Strong	HEAVY	37.5	46.2	41.4	33.3	33.3	33.3	8.0
Big	STRONG	50.0	53.8	51.7	36.1	38.9	37.5	14.2
Heavy	STRONG	56.2	76.9	65.5	41.7	47.2	44.4	21.1
	N[b]	16	13	28	18	18	36	

[a] The number of cases in this cell is reduced by one due to loss of data.

[b] The *S*s reported were those who correctly differentiated the stimulus attribute for all three attributes on all materials.

TABLE 2
PROPORTION ASCRIBING CORRELATED ATTRIBUTE DIFFERENCES TO FACES

Stimulus Differences	Response	First Grade			Sixth Grade			
		Boys	Girls	Total	Boys	Girls	Total	Differences
Happy	CLEAN	12.5	5.9	9.1	0.0	11.1	5.6	3.5
Pretty	CLEAN	28.6	7.7	18.5	5.6	5.6	5.6	12.9
Clean	HAPPY	37.5	47.0	42.4	5.6	22.2	13.9	28.5*
Pretty	HAPPY	85.7	91.7	88.5	67.1	38.9	50.0	38.5*
Clean	PRETTY	100.0	82.4	90.9	50.0	50.0	50.0	40.9*
Happy	PRETTY	81.2	82.4	81.8	83.3	66.7	75.0	6.8
Clean	GOOD	50.0	70.6	60.6	27.8	33.3	30.6	30.0*
Happy	GOOD	100.0	94.1	97.0	77.8	55.6	66.7	30.3*
Pretty	GOOD	78.6	50.0	65.4	44.4	22.2	33.3	32.0*
	N[a]	16	17	33	18	18	36	

[a] Changes in *N* due to failure to name stimulus difference correctly: first-grade boys for PRETTY stimulus 14, girls 12; sixth-grade boys, 17, and girls, 16.
* Significant at the .01 level.

DISCUSSION

The slight rate of change with respect to discrimination of physical dimensions suggests the kind of learning to be expected where the criteria are most obvious.

With the personal attributes the findings are both more extreme and more uneven. It is clear that the only term of those used which designates a simple visible trait—CLEAN—is the one seldom offered to describe any other attribute. This finding agrees with the fact that BIG, the most obvious physical attribute, was less often used with the physical materials to describe other differences.

The other evaluative dimensions refer to more complex traits that are not entirely logically independent. PRETTY may also be said to designate a physical characteristic, but one that adults use both as a constant and a temporary trait, so that a clean smiling face might be deemed prettier than a dirty or frowning one. Thus, while the traits are discriminable they are not independent in the sense that the physical traits are.

HAPPY was used to designate a smiling face, but as children learn the use of the term they may find that it refers to a state of feeling only partially correlated with external evidence. Thus even the smiling–nonsmiling distinction might bear only an imperfect relation to the term HAPPY. One of the largest changes was in the use of HAPPY in describing the prettier girl, and the drop was most extreme in girls—a drop of 53%. The children were from the start, especially the boys, only moderately likely to describe the cleaner girl as HAPPIER.

One possible reason for a drop in the ascription of terms to a correlated difference is a change in metaphorical treatment of pictures. Occasionally children would refuse to say that both children were the same in hair color, but said that one had red and one black hair. The hair in the drawing was white, that is, not filled in in the black-and-white outline picture. Younger children, used to story-book imagination, may be less literal about what is on the page. While this might account for the increase in *same* responses with age, it does not predict the direction of the ascription, by the first graders, which was in no case in the opposite direction from that predicted.

Does the failure to differentiate on this test imply that the children use the terms as virtual synonyms for an undifferentiated referent? With respect to the physical dimensions, only one child was so extreme as to use the same terms interchangeably for all three attributes on all the materials. But if we examine the faces test and omit CLEAN which seems to fall out of the pattern, it appears that 62% of the children used HAPPY-GOOD-PRETTY synonymously, in the first grade.

We would like to argue that the factors that appear as clusters of correlated terms in the semantic differential studies of adults derive from empirical correlations of attributes. They could, of course, be linked purely by verbal associations, as in "He's a good clean player."

If this were the case, differentiation of reference might still be accompanied by semantic differential correlations, since many of the terms on the semantic differential can be applied to the "concept" only metaphorically. DEMOCRACY is clean, fragrant, and sweet only in the poetic sense.

While verbal associations may be one source of such dimensions, we are proposing that the history of concept development in the child provides another source. What remains as a connotative, metaphorical relationship in adults may in many cases start as denotative nondifferentiation. In a sense, the child might be said to acquire first a concept, for instance, of "big-strong-heavy . . ." in other words, a potency referent. The terms he applies to this referent may variously be BIG, STRONG, HEAVY. He may prefer one of these terms for people, another for boats, another for baseballs. Presumably he will only come to differentiate the terms and apply them appropriately to different stimulus dimensions when uncorrelated instances occur and he is corrected, or hears others differentiate the terms. By chance, the sample he selects may have a 100% correlation and he may not encounter errors immediately.

SUMMARY

The physical dimensions of size, weight, and strength are empirically correlated. If the correlation delays discrimination of these attributes as referents for descriptive terms, then younger children should more often use incorrect terms to describe differences between objects. The terms GOOD, PRETTY, CLEAN, and HAPPY should also be used as synonyms prior to differentiation.

A set of materials was prepared in which size, weight, and strength were independently varied in pairs of objects. First-grade children more often than sixth graders said that the pairs of objects differed on other dimensions in addition to the attribute actually contrasted. In a set of pictures of faces, over half of the youngest children treated GOOD, PRETTY, and HAPPY as interchangeable synonyms. The proportion dropped markedly with age. The more easily identified traits, such as the referents of BIG and CLEAN, were least often confused with other attributes.

The results are interpreted as showing that attributes which have metaphorical and connotative links in adult usage, may be denotatively confused at first. The factors found by Osgood on the semantic differential studies of verbal meaning may actually be the referents for several terms used as synonyms, prior to differentiation of finer distinctions between attributes.

REFERENCES

BRUNSWIK, E. *Systematic and representative design of psychological experiments.* Berkeley: Univer. California Press, 1947.

LEOPOLD, W. F. *The speech development of a bilingual child.* Evanston, Ill.: Northwestern Univer. Press, 1939–50.

OSGOOD, C. E., SUCI, G. J., & TANNENBAUM, P. H. *The measurement of meaning.* Urbana: Univer. Illinois Press, 1957.

(Received August 7, 1959)

Journal of Abnormal and Social Psychology
1960, Vol. 61, No. 2, 276-286

A THERAPEUTIC APPROACH TO SCHIZOPHRENICS OF EXTREME PATHOLOGY:

AN OPERANT-INTERPERSONAL METHOD[1]

GERALD F. KING
Michigan State University
STEWART G. ARMITAGE
Veterans Administration Hospital, Battle Creek, Michigan
AND JOHN R. TILTON
System Development Corporation

ONE of the most severe forms of psychopathology is seen in cases culminating in a "vegetative" level of adjustment characterized by the patients being extremely withdrawn, anergic, and uncommunicative. Standard treatment procedures do not produce any appreciable benefits. It would seem that different, if not novel, therapeutic approaches need to be explored if changes are to be effected with this type of patient. This paper presents an evaluation of a therapeutic method that was devised for chronic schizophrenics with such extreme withdrawal symptoms.

Given the label "operant-interpersonal," the method of therapy that is proposed stems from recent work employing learning techniques with schizophrenics. Peters and Jenkins (1954) attempted to retrain chronic schizophrenics by using first simple psychomotor tasks and then increasing their complexity. To increase motivation, subcoma injections of insulin were given with fudge as a reward. Tilton (1956), employing instrumental motor and verbal techniques with even more debilitated patients, found that candy was an adequate reward in itself. His results revealed that mute patients responded better to the experimental sessions (mainly psychomotor in nature) than more verbal patients, while just the opposite trend was observed for "standard" verbal therapy, one of his control conditions. In a similar vein, Lindsley (1954) has discussed the possible extension of operant conditioning techniques to therapy.

OPERANT-INTERPERSONAL METHOD: RATIONALE

The therapeutic goal is to elicit more interest in the environment and to promote more appropriate response patterns. Assumptions about the underlying nature of schizophrenic behavior are not reflected in the therapeutic procedure, as virtually no attention is given to the patient's previous experiences or pathology. Thus, in light of the common tools associated with psychotherapy (Shoben, 1949), the operant-interpersonal method might be considered an unsophisticated therapeutic approach. Be that as it may, the following considerations are offered as the basis for developing the present therapeutic procedure with *highly withdrawn* patients:

1. The therapeutic sequence should be built on responses that are in the repertory of the patient. Verbal communication is an inappropriate response medium because it has all but been abandoned by many patients as part of the withdrawal pattern. Ward observations reveal that although such patients show little verbal activity, they still shuffle about the ward, manipulate eating utensils, smoke cigarettes, and so forth. In short, motor responses are more intact and, hence, more usable.

2. The course of therapy should allow an easy transition from the simple to the more complex. An increase in complexity refers to the addition of verbal and interpersonal components to the therapeutic structure.

3. In the role of "friendly leader," the therapist

[1] Originally conceived by the senior author and David W. Merrell, this research was conducted at Veterans Administration Hospital, Battle Creek, Michigan, during 1956.

The authors are indebted to the following individuals for assistance: Carl R. Brown, being largely responsible for the design and construction of the apparatus; Norman Graff and David W. Merrell, for collecting portions of the data; M. Ray Denny and Norman R. F. Maier, for their interest and suggestions; the late Edmund D. Barry, for conducting the recreational therapy group; Terrence M. Allen, for his statistical consultation; and Frank Restle, for his advice on the preparation of this manuscript. Grateful acknowledgement is also due John Berghost and other personnel of the Medical and Nursing Services for their extensive cooperation.

should provide guidance in a social atmosphere that is conducive to learning. The therapist should also function as a decision maker, continually evaluating the patient's behavior and judging whether or not the patient is ready to cope with more complex situations.

METHOD

Measurements

The following instruments constitute the principal measuring devices that were employed. Others of a minor nature will be described in the analysis of the data.

Indices of status. In earlier work with deteriorated, uncommunicative patients, two of the authors devised the Extreme Mental Illness Schedule (EMIS), an interview procedure for assessing low levels of adjustment. A similar scale, which was apparently developed at approximately the same time, has been reported by Farina, Arenberg, and Guskin (1957). The EMIS consists of 22 variables, which are rated in an interview setting on well defined four- or five-point scales. The first 10 variables involve ratings on what might be called "simple interpersonal tasks" (e.g., response to handshake, response to offer of a chair, response to offer of a cigarette). The remaining variables were assembled from other rating scales, the emphasis being on nonverbal behavior (e.g., posture, psychomotor tempo, peculiar gestures). The sum of the ratings on all the scales yields a morbidity score, with higher scores representing greater severity of neuropsychiatric illness.

Using two staff psychologists to make independent EMIS ratings resulted in an interrater correlation of .95 ($N = 62$). With two different judges and different patients, a correlation of .92 was obtained for interrater reliability ($N = 50$). When the EMIS was readministered to 33 patients (all having achieved stable adjustments) after an interval of 8 weeks, the test-retest reliability was found to be .85. Several psychologists and psychiatrists were asked to rank small groups of patients on severity of illness. Converting the EMIS scores to ranks yielded validity coefficients (rho's) that ranged from .62 to .98. It was also possible to discriminate significantly between chronic schizophrenics with no privileges and chronic schizophrenics having limited privileges on the basis of EMIS morbidity scores.

Clinical status, as reflected by ward adjustment, was assessed with the Ward Observations Scale (WOS), the final 22 items of the Multidimensional Scale for Rating Psychiatric Patients (Lorr, 1953). In a pilot study, two attendants rated 20 chronic schizophrenics on the WOS, resulting in a correlation of .92 between the two independent morbidity scores. When a different attendant rated the same patients after an interval of 2 weeks, the test-retest correlation was .86.

A five-point scale for measuring amount of verbalization (AVS) was designed to be filled out at the end of the EMIS interview. The independent AVS ratings made by two staff psychologists correlated to the extent of .89 ($N = 62$). After 8 weeks, the test-retest reliability was .81 ($N = 31$).

Measures of clinical improvement. Difference scores obtained from successive ratings on the EMIS, WOS, and (to some extent) AVS provide measures of clinical improvement. In addition, the Clinical Improvement Scale (CIS) (King, Merrell, Lovinger, & Denny, 1957) was employed. The first section of the CIS consists of a simple five-point scale, in which the anchoring descriptions range from "has become worse" to "shows very marked improvement." In the event clinical improvement of any degree is indicated, the second section provides a check list which allows the rater to specify the nature of the behavioral changes. Previous studies (King, et al., 1957; Tilton, 1956) have indicated that ward personnel can make reliable assessments on the global scale of the CIS (first section). The following intercorrelations among the principal measures of clinical improvement were obtained ($N = 50$): EMIS vs. WOS, .51; EMIS vs. CIS, .62; and WOS vs. CIS, .45.

Subjects

Selection. All *Ss* were drawn from a large locked ward which housed approximately 300 long-term, male psychotics. The following criteria were used in soliciting possible *Ss* from the ward personnel: diagnosis of schizophrenia, under 45 years of age, low verbal and physical activity level, and no major change in behavior for at least two years. This procedure yielded 92 patients who were screened with interviews and then administered the EMIS and AVS. With the latter instrument, all patients were classified for level of verbalization using the following categories: mute or nearly mute, minimal verbalization, and below normal verbalization. Four groups of 12 *Ss* each were formed so that each group contained four *Ss* from each of the three verbalization levels. The *Ss* of the four groups were individually matched on severity of illness (EMIS) and length of hospitalization.

Characteristics. The ages of the final 48 *Ss* ranged from 25 to 42 years ($M = 33.8$, $SD = 4.4$), with length of hospitalization extending from 50 to 171 months ($M = 108.0$, $SD = 33.3$). Level of education was from 7 to 13 years ($M = 10.1$, $SD = 1.9$). The distribution of the schizophrenic subtypes was as follows: hebrephrenic, 18; undifferentiated, 14; catatonic, 11; and paranoid, 5. No significant differences existed among the four groups for these variables. All 48 *Ss* had in the past undergone at least three of the following treatment procedures: electroshock therapy, insulin coma therapy, ataractic drug therapy, psychotherapy, and group therapy.

The Operant-Interpersonal Method

Apparatus. Section A of Figure 1 provides a diagram of the Multiple Operant Problem-Solving Apparatus (MOPSA), which was constructed to put the operant-interpersonal method into practice. A modification of the apparatus used by Skinner, Solomon, and Lindsley (1955–56), MOPSA was built into an 8 by 8 ft. panel erected across the alcove of a large room. Projecting from the middle of the panel, 35 in. from the floor, were three wooden levers (6 in. in length) arranged in a row and spaced 5 in. apart. Below the outer levers were slots opening into trays, which were used for dispensing

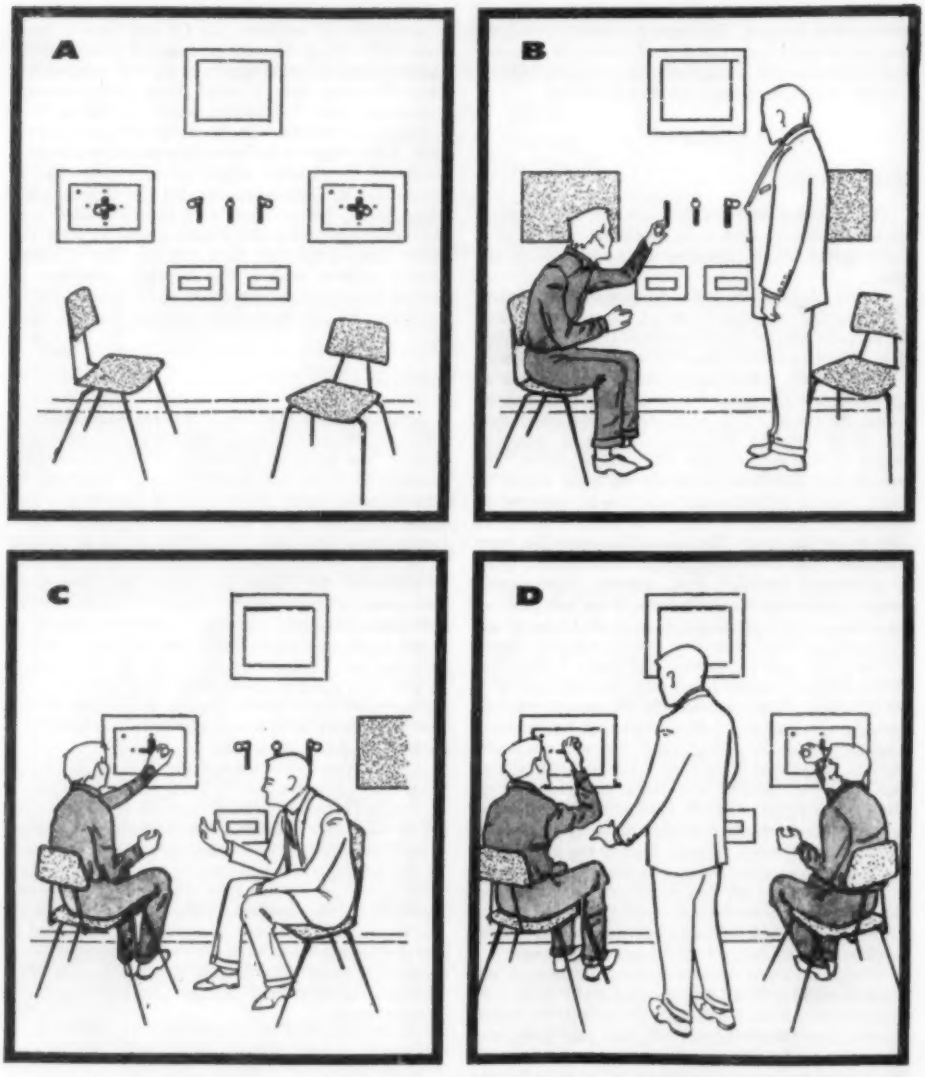

FIG. 1. The operant-interpersonal method: apparatus and illustrations of the therapeutic phases. (A. Multiple Operant Problem-Solving Apparatus—MOPSA, B. Simple Operant Behavior—Phase I, C. Simple and Complex Problem-Solving—Phase II, D. Cooperative Problem-Solving—Phase III)

rewards of candy and cigarettes. A 14 by 14 in. screen for projecting slides was located above the middle lever protected by a Plexiglas cover. Each lever could be depressed 3 in. against a force of 3 lbs., as regulated by an extension spring. In order to obtain rewards (candy, pictures, cigarettes), it was necessary for *S* to depress the appropriate lever until there was a "click" followed by the ringing of a chime.

Lateral to each of the outer levers by 8 in. were two 11 by 13 in. aluminum plates, which could be concealed by covers. A cross, into which a lever could be inserted, had been cut in the center of each plate. Counterbalanced by just enough tension to return and maintain them at the cross intersections, the levers could be moved freely in the two planes of the crosses. If *S* moved the lever of either cross from the position of rest

to the end of the correct crossarm, a green light at the end of the arm became lit. Through a system of relays, problems could be introduced requiring from one to four movements (lights), along with various patterns of the "right-left-up-down" sequence. Movements had to be executed in a prescribed order so that an error canceled any previously correct responses, requiring that the entire sequence be started again. In addition to the green lights at the crossarms, an amber light located at an upper corner of each plate flashed when a problem was solved. By changing the relay system, new problems could be introduced in a few seconds. As the two crosses could be operated independently or together, problems requiring cooperation between *Ss* could be employed.

Temporary screens enclosing an 8 by 10 ft. area were erected in front of MOPSA to reduce distraction and to restrict the movements of *S*. During the initial stages of therapy, the only furnishings in this area were two chairs. Behind the panel and not visible to *S*, a therapy-assistant operated the projector (scenic slides), released the candy (Hershey chocolate kisses) and cigarettes (Lucky Strikes), operated the relay system, and recorded the readings of the counters attached to the operant levers.

Therapeutic procedure. The 12 *Ss* undergoing operant-interpersonal therapy were seen three times a week for 3½ months (15 weeks). Therapeutic sessions lasted from 20 to 30 minutes. In adapting the operant-interpersonal method to the individual *S*, variations were bound to occur. However, the therapeutic procedure can be divided into the following three phases which applied to all *Ss*: Phase I, Simple Operant Behavior; Phase II, Simple and Complex Problem-Solving; and Phase III, Cooperative Problem-Solving.

At the beginning of Phase I, each *S* was led into the therapy room and seated in front of MOPSA. After directing *S*'s attention to the three operant levers, the therapist briefly described their operation. The explanation was followed by a demonstration in which the therapist manipulated each of the levers, pausing to present the candy and cigarettes to *S* and to call his attention to the picture on the screen. The *S* was then urged to depress one of the levers. In a majority of the cases, it was necessary to repeat the demonstration. In a few cases, as a last resort, the therapist placed *S*'s hand on the lever and then supplied the necessary force for depressing the lever. With one *S*, the therapist initially unwrapped the candy reward and placed it in the *S*'s mouth. If *S* successfuly manipulated a lever, his performance was greeted with a brief but definite acknowledgment (e.g., "good," "very good," "good work," etc.). Section B of Figure 1 gives an illustration of this simple operant behavior. Verbalization by the therapist was held to a minimum, and what was said was mainly directed toward the operation of MOPSA.

Although there was marked variability among *Ss*, all eventually began to respond voluntarily. When the individual *S* reached this point, frustration was introduced in the form of nonreinforced trials. When any of the *Ss* stopped pressing the levers (extinction) after nonrewarded trials, special attention (instructions and demonstrations) was given by the therapist until *S* was willing and able to respond to various conditions of nonreinforcement. Phase I lasted approximately 4 weeks.

In Phase II, one of the aluminum plates was uncovered, and a lever was inserted into the cross. Beginning with the simplest type of problem solving, the therapist demonstrated the one-light problem, which required the lever to be moved into a single crossarm. He showed that only one of the four possible movements turned on a green light (at the end of the crossarm) and subsequently the amber light (upper corner of the plate). After making the correct response on the cross, the therapist depressed one of the operant levers and called *S*'s attention to the resulting reward. When *S* could satisfactorily solve a variety of one-light problems, he was gradually introduced to problems requiring more movements. Eventually all *Ss* reached at least the level of two-light problems, while five of the *Ss* consistently solved four-light problems. This problem solving behavior is illustrated in Section C of Figure 1.

After approximately 3 weeks, verbalization was integrated into the problem-solving process. For example, *S*, after solving a two-light problem several times, might be stopped after making the first movement and asked, "Now where do you go next?" If the therapist failed to elicit a verbal response from a mute *S*, he would ask the *S* to point to the next movement. Another procedure was to ask *S*, after he had solved a problem and received a reward, to describe all the movements required in the problem. When *S* was confronted with a new problem, he might be asked, "What is going to be your first move?" At this time, the therapist also encouraged more verbal interaction at the beginning and close of sessions, with topics not directly related to MOPSA being discussed (e.g., "What did you have for breakfast this morning?"). Phase II lasted about 6 weeks.

The second cross was uncovered at the outset of Phase III, which was devoted to cooperative problem solving. Seated in front of the second cross, the therapist informed *S* that they would have to work as a team in order to obtain rewards. In a simple two-light problem, the therapist made the first movement and *S* the second, as well as vice versa. Quality of performance seemed to drop sharply in this interpersonal context, as most *Ss* experienced difficulty in coordinating their efforts with another individual. As progress did occur, the therapist and *S* proceeded to three- and four-light problems in which all possible therapist-subject orders of responding were required. As in the latter stages of Phase II, verbal interaction was emphasized.

When all *Ss* had considerable experience working with the therapist, two *Ss* functioned as a team. This feature of Phase III is shown in Section D of Figure 1. Then, two additional *Ss* entered the therapeutic situation, standing behind and observing the cooperating *Ss* before being called on to give either motor or verbal solutions. Stand-by chairs were pulled into the therapeutic area, and eventually as many as six *Ss* attended a single session. In this way, it was possible for every *S* to cooperate with every other *S* within a short period of time. Occasionally, the therapist again functioned with more advanced *Ss* and purposely made mistakes, placing the burden of making the corrections on the *Ss*. A by-product of this procedure was the introduction of some mild humor into this situation. Problems were also devised in which two *Ss*, after performing successfully, were required to communicate the solution to

naive *S*s (nonobservers). Phase III was terminated after about 5 weeks.

Control Groups

The following provides brief descriptions of the control groups which were conducted concurrently with the operant-interpersonal group.

Verbal therapy. The 12 *S*s undergoing verbal therapy were given the same amount of time as the operant-interpersonal *S*s, three sessions of 20 to 30 minutes per week. Therapy was adapted to the individual *S*. The therapist made every effort to establish verbal communication with them on whatever basis was possible. Depending on the *S*, a variety of topics were explored, ranging from routine daily events to *S*'s feelings and concerns. With mute *S*s, the sessions consisted of monologues by the therapist, punctuated by hopeful pauses. When operant-interpersonal therapy began using more than one *S*, a similar procedure was followed in verbal therapy. Hence, in the latter stages verbal therapy became group verbal therapy.

Recreational therapy. The 12 *S*s in recreational therapy participated in a variety of activities, usually within a larger group of patients. Time spent in this form of therapy ranged from 3 to 5 hours per week.[2] The following indicates the nature of some of the activities: medicine ball manipulation, shuffleboard, rhythm band, singing, social games (e.g., hokey-pokey), dancing with volunteer workers. For the purpose of the present research, the *S*s were given more individual attention than they ordinarily would have received.

No therapy. Although the 12 *S*s in this group were not given any special treatment, they were allowed to continue to participate in the ward activity schedule (e.g., occupational therapy, corrective therapy), as did the other 36 *S*s on a more limited basis. Heretofore, the program had produced no discernible change in the patients of the ward. Except when it was not possible (e.g., swimming), most of the *S*s slept during the activities.

Procedure

All 48 *S*s were transferred to a single dayroom in which they were the sole occupants. A specially selected attendant was transferred from night duty to function as an attendant-observer on the dayroom for several hours each day.

The *S*s had been administered the EMIS and AVS by one of the authors prior to their being assigned to groups. Teams of attendants and nurses rated the *S*s on the WOS. A staff psychologist not previously associated with the research readministered the EMIS and AVS at the end of the project (15 weeks later). The attendant-observer rated the *S*s again on the WOS and filled out the CIS. Although the attendant-observer was aware that the *S*s were participating in a special project, he was not acquainted with the nature of the various groups or their memberships.

The *S*s of the operant-interpersonal and the verbal therapy groups were divided among two therapists (two of the authors) who were assigned six *S*s from each group. Each of the four subgroups were represented by two *S*s from each level of verbalization. The decision to use the two available therapists in this manner might be questioned since the same therapists saw the *S*s in both the operant-interpersonal group and one of the control groups. Although the personality characteristics of the therapists would be held constant by this procedure, there was the possibility of differential effort or interest. To combat any possible bias, the therapists attempted to consider each *S* on an individual basis and to utilize all possible resources within the requirements of the research design. Detailed notes were kept for all therapeutic contacts. The therapists periodically monitored each other's sessions as hidden observers. The *S*s receiving recreational therapy were assigned to an experienced activity-therapist who was particularly interested in chronic schizophrenics.

RESULTS

Since the *S*s from the operant-interpersonal and verbal therapy groups were divided among two therapists, a preliminary analysis was concerned with any differences in the two groups attributable to the therapists. Improvement scores derived from the EMIS and WOS were each treated by analysis of variance. As none of the relevant F ratios were statistically significant at the .05 level, the null hypothesis was accepted, i.e., no differences produced by the therapists.

Differential Clinical Improvement

Table 1 gives the results of the analysis of variance for the EMIS improvement scores. As can be seen, a significant difference among the therapeutic groups was indicated by an F ratio of 7.313 ($p < .001$). Initial status (severity of illness and/or length of hospitalization) also produced a significant F ratio ($p < .05$) using

TABLE 1
ANALYSIS OF VARIANCE OF EMIS IMPROVEMENT SCORES

Source	*df*	*MS*	*F*
Therapeutic Groups (T)	3	58.965	7.313**
Level of Verbalization (V)	2	22.312	1.057
Initial Status within V (I)	9	21.118	2.619*
T × V	6	3.924	0.487
T × I	27	8.063	
Total	47		

* $p < .05$.
** $p < .001$.

[2] The group receiving recreational therapy was a somewhat unusual control group. Due to a misunderstanding, this group was seen more than twice as much as the operant-interpersonal and verbal therapy groups.

T × I as the error term, which is not entirely appropriate in this case. Contrary to Tilton's (1956) findings, no significant difference was found for level of verbalization, although it should be pointed out that only *S*s relatively low in verbalization were used in the present research. The operant-interpersonal group showed the most EMIS improvement, with the control groups being in the following descending order: recreational therapy, no therapy, verbal therapy. It was considered appropriate, and not a post hoc analysis, to compute *t* tests between the operant-interpersonal group and the means of each of the control groups. The results, using two-tailed tests, revealed that the operant-interpersonal group improved significantly more than the recreational therapy group at the .02 level and both the no therapy and verbal therapy groups at beyond the .01 level.

When the WOS improvement scores were treated by analysis of variance, only the difference among the therapeutic groups was significant ($F = 4.652$, $p < .01$). In fairly close agreement with the previous analysis, the descending order of the means of the groups on improvement was as follows: operant-interpersonal method, recreational therapy, verbal therapy, no therapy. While *t* tests revealed that the difference between the operant-interpersonal and recreational therapy groups was not statistically significant, the operant-interpersonal group showed a significantly higher level of improvement than the verbal therapy and no therapy groups, both at the .01 level. The full Lorr scale consists of 12 factors or dimensions derived from factor analysis, with the WOS contributing at least three items to each of the following factors: activity level, compliance vs. resistiveness, submissiveness vs. belligerence, and withdrawal. Using scores based on each of the four factors, differences among the therapeutic groups were further explored with analysis of variance. Only improvement in the area of withdrawal yielded a significant difference ($F = 3.827$, $p < .05$).

TABLE 2

COMPARISON OF THE THERAPEUTIC GROUPS: DISTRIBUTION OF THE SUBJECTS ON THE CLINICAL IMPROVEMENT SCALE (CIS)

CIS[a] Ratings	Groups			
	Operant-Interpersonal	Verbal Therapy	Recreational Therapy	No Therapy
Has become worse	0	1	0	0
Essentially no change	4	10	8	10
Shows minor improvement	3	1	2	2
Shows considerable improvement	5	0	2	0

[a] The rating "shows very marked improvement" is not included in the table since no *S* was rated in this category.

TABLE 3

COMPARISON OF THE THERAPEUTIC GROUPS ON CHANGES IN LEVEL OF VERBALIZATION

Changes in Level of Verbalization (AVS)	Groups[a]			
	Operant-Interpersonal	Verbal Therapy	Recreational Therapy	No Therapy
Decreased	0	6 (7)	2	1
Unchanged	5 (5)	6 (5)	8	9
Increased	7 (7)	0	2	2

[a] The numbers in parentheses for the operant-interpersonal and verbal therapy groups represent appraisals of changes in level of verbalization obtained from the notes kept by the therapists.

Further support for the greater effectiveness of operant-interpersonal therapy can be found in Table 2, which presents a comparison of the therapeutic groups on the CIS. More *S*s in the operant-interpersonal group (eight) exhibited some degree of clinical improvement than *S*s in any of the control groups. Ranking the groups on amount of improvement from Table 2 results in the identical order as that obtained for the EMIS scores. The operant-interpersonal group was compared with each of the control groups with the Fisher exact test by using the dichotomy "improvement vs. no improvement" and constructing fourfold tables. With this analysis, the operant-interpersonal group was significantly better than the verbal therapy ($p = .009$) and no therapy ($p = .019$) groups; the trend in favor of the operant-interpersonal over the recreational therapy group was not significant ($p = .220$). Although nothing is known about its reliability, the CIS checklist (second section) provides information on the qualitative aspects of clinical improvement. An inspection of the checklist indicated that the nature of the clinical changes was quite similar for the improved *S*s of the operant-interpersonal and control groups. The largest discrepancies occurred on the checklist items "improved

verbal communication," where proportionately more improved *S*s of the operant-interpersonal group were checked, and "less tension," where the incidence of controls was higher.

Changes in level of verbalization, as obtained from successive administrations of the AVS, are given in Table 3. The operant-interpersonal group was again superior. More *S*s in this group showed an increase in verbalization, and it was the only group in which no *S* dropped in amount of verbalization. A striking feature revealed by Table 3 is that at least six *S*s in the verbal therapy group showed a decrease in verbalization. When the dichotomy "increase vs. no increase" was used for the verbalization changes, the Fisher exact test revealed that the operant-interpersonal group was significantly different from the verbal therapy group ($p = .005$). The comparisons of the operant-interpersonal group with the recreational therapy and no therapy groups both approached significance ($p = .089$). These tests of significance are conservative ones, since the classification for change does not take into consideration decreases in verbalization.

During the last 3 weeks of the therapeutic sessions, one of the authors began attending as an observer the meetings of the ward personnel in which various patients were discussed and evaluated. At this time, patients were being considered for transfer to a small "privilege" section which had been opened on the ward. The observer recorded the names of all patients receiving favorable consideration for transfer. None of the ward personnel participating in the meetings knew what treatment procedures were being given the patients who were serving as *S*s in the research. Listed according to group memberships, the number of *S*s recommended for transfer was as follows: 5 (operant-interpersonal), 0 (verbal therapy), 2 (recreational therapy), 0 (no therapy). Actually, only one *S*, a member of the operant-interpersonal group, was transferred to the open section within 4 weeks after termination of the therapeutic sessions.

Subsequent evaluations. Some data on the status of the therapeutic groups were collected 6 months after the therapeutic sessions were discontinued. At this time, a team of ward personnel rated the *S*s of the operant-interpersonal, verbal therapy, and no therapy groups on the WOS. The recreational therapy group was not included since it, as an exception, was still receiving therapeutic attention. Treatment of the WOS morbidity scores by analysis of variance yielded a significant difference among the three therapeutic groups ($F = 4.239$, $p < .05$), with the operant-interpersonal group having the least pathology. The operant-interpersonal group was the only one showing a significant improvement over the initial WOS rating ($t = 2.664$, 11 *df*, $p < .05$). The transferred *S* from the operant-interpersonal group was still functioning on the open section. However, no other *S*s had been transferred during this time interval.

Other Comparative Data

When the screening interview was conducted, none of the *S*s, as best as it could be determined, indicated any interest in leaving the ward for any purpose. During the thirteenth week of the therapeutic sessions, a staff psychologist not previously associated with the research briefly interviewed all *S*s of the operant-interpersonal, verbal therapy, and no therapy groups. Each *S* was rated for motivation to leave the ward (either to an open ward or out of the hospital) on the following scale: 1—no interest, 2—some interest, 3—definite interest. It can be seen in Table 4 that more *S*s in the operant-interpersonal group displayed some degree of interest in leaving the ward than the *S*s in the two control groups. When this measure of motivation was divided into "interest vs. no interest," the Fisher exact test indicated that these differences were significant ($p = .019$) when the operant-interpersonal group was compared with each of the control groups.

As was noted earlier, the *S*s of the no therapy

TABLE 4

COMPARISON OF THE THERAPEUTIC GROUPS ON DEVELOPMENT OF MOTIVATION TO LEAVE THE WARD

Ratings on Motivation to Leave the Ward	Groups		
	Operant-Interpersonal	Verbal Therapy	No Therapy
No interest	4	10	10
Some interest	5	1	2
Definite interest	3	1	0

group continued to participate in the ward activity schedule. Although less time was available due to the various therapeutic procedures, the other *S*s also attended these activities. The one activity in which all *S*s participated was occupational therapy. At the close of the therapeutic sessions, the occupational therapist was given a list of all *S*s and asked to indicate any who had displayed "increased interest or more constructive behavior" in occupational therapy during the preceding 2 months. The resulting distribution of *S*s rated as showing some positive change was as follows: 4 (operant-interpersonal), 0 (verbal therapy), 1 (recreational therapy), 0 (no therapy).

In the initial evaluation of the *S*s, it was found that each therapeutic group contained from one to three *S*s classified as enuretic. At the close of the therapeutic sessions, only one *S* was designated as improved by the night attendants, and this *S* was a member of the operant-interpersonal group.

If the therapy notes compiled by the therapists are used, another comparison between the operant-interpersonal and verbal therapy groups can be made. The notes indicated that during the first 3 weeks two *S*s from each group showed resistance, i.e., resistance to leaving the ward for the therapeutic sessions. No *S* from the operant-interpersonal group behaved in this manner during the last 3 weeks, but five from the verbal therapy group were described as resistant.

Clinical Observations

The progress shown by the *S*s in the operant-interpersonal group was readily discernible to the authors and other daily observers. To give an example, there was the day during the eleventh week of therapy that a long-term mute *S* surprised the therapist by greeting him with, "Good morning. Got a match?" The authors were particularly impressed with the interest value that the therapeutic tasks held for the *S*s. More than a month after therapy had been terminated, an apparently puzzled attendant contacted one of the authors to inquire about the "lights," as two former *S*s were constantly asking him when they could resume this activity. Observations of the *S*s in the verbal therapy group paralleled the formal analysis. A number of *S*s in this group obviously became worse. Strong negativism toward the therapeutic sessions was common, and practically all of the eight verbal *S*s responded less, several becoming mute.

Interrelationships among Subject, Therapeutic, and Outcome Variables

The interrelationships among the following types of variables were explored with the 12 *S*s of the operant-interpersonal group: subject, therapeutic, and outcome. The subject variables were age, length of hospitalization, severity of illness, and level of prepsychotic adjustment. Severity of illness was indicated by a composite of the EMIS and the WOS. Using a scale derived partly from certain prognostic criteria (Kantor, Wallner, & Winder, 1953; Phillips, 1953), one of the authors ranked the *S*s on level of prepsychotic adjustment on the basis of case history information. Previous research (King, 1958) has shown a relatively high interrater reliability for ranking a similar dimension from case history material. The therapeutic notes were used by two of the authors who collaborated in ranking the *S*s on the following therapeutic variables: operant rate, level of problem solving, and task improvement. Rate of operant response (number of responses) was based on performance during the first 4 weeks of therapy. Level of problem solving pertained to the ability of the *S*s to solve the light-sequence problems of Phase II. Task improvement reflected the ranking of the *S*s on overall improvement in coping with the task requirements throughout the course of therapy. Outcome was represented by the relative standings of the *S*s on clinical improvement, as indicated by a composite of the improvement scores from the EMIS, WOS, and CIS.

The rank-order intercorrelations among the various subject, therapeutic, and outcome variables can be examined in Table 5. Of the subject variables, only severity of illness was related to any of the therapeutic variables, the rho's with operant rate and level of problem solving being significant at the .05 and .02 levels, respectively. The negative correlation between severity of illness and operant rate coincides with the results reported by Mednick and Lindsley (1958) with chronic patients. Although there was a trend for length of hospi-

TABLE 5
RANK-ORDER INTERCORRELATIONS AMONG SUBJECT, THERAPEUTIC, AND OUTCOME VARIABLES
($N = 12$)

Variables	2	3	4	5	6	7	8
1 Age	.804****	−.139	.028	−.070	.028	−.244	−.434
2 Length of Hospitalization		.189	.140	−.301	−.322	−.378	−.545*
3 Severity of Illness			.105	−.601**	−.685***	−.566*	−.266
4 Level of Prepsychotic Adjustment				.119	.098	.091	−.580**
5 Operant Rate					.699***	.615**	.203
6 Level of Problem Solving						.769****	.294
7 Task Improvement							.594**
8 Clinical Improvement							

Values of p (two-tailed tests):
* $p < .10$.
** $p < .05$.
*** $p < .02$.
**** $p < .01$.

talization to be negatively correlated with outcome ($p < .10$), the one subject variable significantly correlated with clinical improvement was level of prepsychotic adjustment ($p < .05$). Since it was negative, this latter correlation would appear to be inconsistent with previous findings which have shown that level of prepsychotic adjustment, as a reflection of more reactive than process characteristics, was positively related to prognosis. A possible explanation lies in viewing this finding in the light of an analysis presented by Zubin and Windle (1954), who found that certain prognostic indices were positively related to favorable outcome in acute psychotics but negatively related for chronic psychotics (the status of the present *S*s). All the therapeutic variables were positively intercorrelated at the .05 level or better, and each correlated positively with outcome. However, only the correlation between task improvement and clinical improvement was significant ($p < .05$). This correlation is compatible with the important assumption that therapeutic gains generalize to behavior outside of therapy. As in most post-mortem discussions, reasonable interpretations on the whole can be posited for the obtained correlations.

DISCUSSION

In giving fairly clear support for the therapeutic effectiveness of the operant-interpersonal method, the results concur with those reported by Peters and Jenkins (1954) and Tilton (1956) in their evaluations of similar therapeutic methods with chronic schizophrenics. There are few basic differences in the therapeutic orientations employed in these studies. While in the present study the analysis of therapeutic gains was based on measures tapping more facets of clinical improvement, the principal difference between this and the previous studies seems to be that the patients selected for therapy, in addition to being more debilitated, represented a more homogeneous group. The selection of patients was guided by the notion that the operant-interpersonal method was especially applicable to extremely withdrawn schizophrenics. It appeared that restricting the patients to the aforesaid type sharpened the differences between the operant-interpersonal method and the control procedures. For example, evidence was obtained indicating that control patients undergoing verbal therapy actually experienced some adverse effects. This finding is compatible with one of the underlying assumptions of the operant-interpersonal method, expressible as follows: in conducting therapy with patients of this nature, verbal communication should not be emphasized, at least, as an initial procedure.

The present study does not throw any appreciable light on what processes or components, among the many in the operant-interpersonal method, are crucial ones in promoting clinical improvement. How much does the relationship that develops between therapist and patient contribute to success? Without getting into reinforcement theory, one might wonder about the operation of the rewards. For example, is the patient more responsive to social reinforcement ("good") or to more tangible rewards (e.g., candy)? Actually, the different

features of the method pose some problems of more general interest, as attested by the following question: does the introduction of an interpersonal component result in a qualitative change in the therapeutic setting, or does it merely represent increased complexity? In regard to the latter issue, there is some evidence indicating that, instead of or in addition to increased complexity, interpersonal components represent a separate dimension of withdrawal, interpersonal-withdrawal as cpposed to thing-withdrawal (King, 1956).

The practical aspects of the findings bring forward other matters of interpretation. The judgments of the ward administrators led to only one patient of the operant-interpersonal group being transferred to an open ward, not an imposing gain. It is not difficult to think of a reasonable defense for this finding. Since the patients were only seen for approximately 22 hours, there is the possibility of additional clinical improvement with a longer period of therapy. Further sessions could be devoted to extending the previous therapeutic pattern, i.e., increasing the complexity of the therapeutic environment. The therapeutic tasks would gradually require more complex discriminations and more conceptual solutions as the problems at the same time became more and more embedded in a verbal and interpersonal context.

On the other hand, there is the possibility that the obtained degree of improvement represents all that can be expected from the operant-interpersonal method. Such an outcome would be compatible with a commonly accepted orientation which implies that a necessary condition for significant gains in *functional* disorders is therapy that penetrates beneath symptoms. The authors are willing to entertain, are even inclined toward, the possibility of limited further improvement, but for different reasons. For one thing, there is the question of whether the patients involved, instead of being clearly functional, do not have contributing organic factors (to be discussed in the next paragraph). In any event, if the results are accepted as they stand, the practical value of the operant-interpersonal method becomes less evident. No ready answer can be offered for the question "What comes next?" The authors resist the temptation to resort to the rationale so frequently associated with the newer psychopharmacological drugs: that the patients, while not "cured," are made accessible to psychotherapy. Whether or not organicity is a factor is one matter, but independently there are doubts about the extent to which psychotherapy can be applied. There is no convincing evidence that psychotherapy is a feasible method for *chronic* psychotics at any stage of illness.

The operant-interpersonal method might be given a final scrutiny in the light of the type of patients for which it was devised. In classifying such patients as being in the terminal stages of process schizophrenia, many clinicians tend to suspect, or even posit, the presence of organic factors either in their etiology or, at least, current functioning. Some evidence can be cited to support this notion (Brackbill, 1956). If this point of view is accepted for the sake of discussion, then psychological therapies for this class of patients would be peripheral approaches, ones that could only operate within certain "neurophysiological limits." Although beneficial results from purely psychological therapies would not be precluded, the implication is that therapeutic gains would be confined to symptomatic ones. Within the available limits, however, specificity of treatment might well be related to the amount of clinical improvement. According to this outlook, the operant-interpersonal method, by focusing on specific symptomatology (withdrawal behavior), becomes a reasonable *psychological* approach. This line of thought is similar to that offered by Jones (1956) in his discussion of therapeutic procedures for "constitutional" disorders.

SUMMARY

The operant-interpersonal method was developed as a therapeutic procedure for chronic schizophrenics with pronounced withdrawal symptomatology. In order to put the method into practice, the Multiple Operant Problem-Solving Apparatus was constructed. Emphasizing motor behavior, the operant-interpersonal method progressed from the simple to the complex. Initially, the patients undergoing this procedure on an individual basis made simple operant responses in the presence of a therapist for rewards of candy, cigarettes, and colored slides. More complex psychomotor, verbal, and interpersonal components were systematically incorporated into the procedure in accordance

with the patient's progress. At its maximum complexity, the therapeutic environment required each patient to communicate with other patients and to enter into cooperative relationships in order to solve problems.

Operant-interpersonal therapy was given to 12 patients, while an equal number of patients concurrently underwent each of the following control conditions: verbal therapy, recreational therapy, no therapy. For the patients in the operant-interpersonal and verbal therapy groups, therapeutic sessions of 20 to 30 minutes were scheduled three times per week for 15 weeks.

The operant-interpersonal method was more effective than all the control methods in promoting clinical improvement, based both on ward observations and interview assessments. Comparisons on the following variables also yielded differences in favor of the operant-interpersonal method: level of verbalization, motivation to leave the ward, resistance to therapy, more interest in occupational therapy, decreased enuresis, and transfers to better wards. The patients undergoing verbal therapy actually became worse in some ways (e.g., verbal withdrawal). Some comparative data were also collected 6 months after the termination of the therapeutic procedures. Several frames of reference, both theoretical and practical, were used in interpreting the results.

REFERENCES

BRACKBILL, G. A. Studies of brain dysfunction in schizophrenia. *Psychol. Bull.*, 1956, **53**, 210–226.

FARINA, A., ARENBERG, D., & GUSKIN, S. A scale for measuring minimal social behavior. *J. consult. Psychol.*, 1957, **21**, 265–268.

JONES, H. G. The application of conditioning and learning techniques to the treatment of a psychiatric patient. *J. abnorm. soc. Psychol.*, 1956, **52**, 414–419.

KANTOR, R. E., WALLNER, J. M., & WINDER, C. L. Process and reactive schizophrenia. *J. consult. Psychol.*, 1953, **17**, 157–162.

KING, G. F. Withdrawal as a dimension of schizophrenia: An exploratory study. *J. clin. Psychol.*, 1956, **12**, 373–375.

KING, G. F. Differential autonomic responsiveness in the process-reactive classification of schizophrenia. *J. abnorm. soc. Psychol.*, 1958, **56**, 160–164.

KING, G. F., MERRELL, D. W., LOVINGER, E., & DENNY M. R. Operant motor behavior in acute schizophrenics. *J. Pers.*, 1957, **25**, 317–326.

LINDSLEY, O. R. Discussant for B. F. Skinner, A new method for the experimental analysis of the behavior of psychotic patients. *J. nerv. ment. Dis.*, 1954, **120**, 403–406.

LORR, M. Multidimensional scale for rating psychiatric patients: Hospital form. *VA tech. Bull.*, 1953, No. 10-507.

MEDNICK, MARTHA T., & Lindsley, O. R. Some clinical correlates of operant behavior. *J. abnorm. soc. Psychol.*, 1958, **57**, 13–16.

PETERS, H. N., & JENKINS, R. L. Improvement of chronic schizophrenic patients with guided problem-solving, motivated by hunger. *Psychiat. Quart. Suppl.*, 1954, **28**, 84–101.

PHILLIPS, L. Case history data and prognosis in schizophrenia. *J. nerv. ment. Dis.*, 1953, **117**, 515–525.

SHOBEN, E. J., JR. Psychotherapy as a problem in learning theory. *Psychol. Bull.*, 1949, **46**, 366–392.

SKINNER, B. F., SOLOMON, H. C., & LINDSLEY, O. R. New techniques of analysis of psychotic behavior. *ONR tech. Rep.*, 1955–56, No. 3. (Contract N5-ori-07662)

TILTON, J. R. The use of instrumental motor and verbal learning techniques in the treatment of chronic schizophrenics. Unpublished doctoral dissertation, Michigan State Univer., 1956.

ZUBIN, J., & WINDLE, C. Psychological prognosis in mental disorders. *J. abnorm. soc. Psychol.*, 1954, **49**, 272–281.

(Received April 13, 1959)

Journal of Abnormal and Social Psychology
1960, Vol. 61, No. 2, 287-291

CASE REPORT

A PSYCHOLOGICAL CASE STUDY OF DEATH DURING PREGNANCY[1]

ANTHONY DAVIDS AND MORTON SILVERMAN
Brown University and Emma Pendleton Bradley Hospital

In the course of a collaborative project investigating perinatal factors in child development, we have studied more than 100 pregnant women. In addition to exhaustive medical observations and obstetrical recordings on these patients, the women have been administered a comprehensive battery of psychological tests. The aim of the psychological assessment is to add to knowledge about emotional factors in pregnancy, and to relate the findings to medical data recorded in the delivery room. This collaborative investigation is planned as a long-range research program, and eventually we hope to relate emotional factors in the mother, during and shortly after pregnancy, to the subsequent intellectual and emotional development of the offspring. The emphasis in this overall research program is on well-controlled experimental design, large samples of patients, and statistical analyses of data. However, it seems worthwhile to report the findings derived from this individual case, which, fortunately, represents a rare occurrence in research contexts.

Brief Case History

The patient was a 39-year-old white woman of the Protestant faith. She had been married 8 years to a man 11 years her senior and had a 6-year-old son. Immediately following delivery of the first child, the patient underwent cholecystectomy (removal of the gall bladder). Before her discharge from the hospital, a metastasizing carcinoid, originating in her duodenum and involving her pancreas, was diagnosed. In the course of major surgery necessitated by this cancerous condition, the pancreas and spleen were removed, and a surgical hitch-up between the small intestine and the stomach was performed. Removal of the pancreas resulted in diabetes and demanded continuous insulin treatment. Shortly after the operation, a subphrenic abscess developed, requiring further surgery to permit drainage. Shortly thereafter, the patient developed intestinal obstructions requiring reoperation. Then a postoperative hernia developed, with separation to the level of the skin, necessitating further major surgery to insert a tantalum screen in the wall of the upper abdomen. In addition to these abdominal difficulties, the patient had a history of chronic otitis media of unknown onset.

After the stomach surgery, the patient did not menstruate for many months. She had been strongly advised not to become pregnant, and when her menstrual cycles began again, she used contraceptives. When she first learned that she was again pregnant in August 1959, she was quite upset, but then decided that she wanted to have the baby. Her surgeon advised against this and recommended a therapeutic interruption or legal abortion. She was referred to an obstetrician for termination of the pregnancy, but in view of her eagerness to have the child, and with consultation from other qualified medical specialists, the obstetrician accepted her as a pregnancy case. The predicted date of delivery was April 20, 1960.

In late November 1959, she was tested by a clinical psychologist. Her referral for this assessment was on a routine basis, the same as other pregnant women who were being seen as research subjects for the National Collaborative Project.

Shortly after the time of the psycological assessment, the patient developed a severe upper respiratory infection with considerable pain in her ears, high temperature, dizziness, and related symptoms. She was admitted to the Providence Lying-In Hospital for treatment of these ailments and was soon transferred to the Chapin Hospital for communicable diseases, diagnosed as a case of meningitis caused by pneumococcus bacteria. A ruptured abscess in the patient's middle ear sent septic emboli to the base of her brain, and she died very suddenly. At the time of her death, there was an audible fetal heart-beat, so a post-mortem Caesarean section was performed in the attempt to save the child's life. The infant girl lived for 18 hours, then died of hyaline membrane disease of the lungs, a condition peculiar to premature babies and the progeny of diabetics.

Having briefly reviewed this background and

[1] This study was made possible by a research grant, No. B-2356, from the National Institute of Neurological Diseases and Blindness, Public Health Service, awarded to the Brown University Institute for Research in the Health Sciences. The present report stems from an ancillary study to the National Collaborative Project, directed locally by Glidden Brooks at Brown University, which is investigating perinatal factors in child development. We wish to express our indebtedness to Bertram H. Buxton, Jr., obstetrician on the Providence Lying-In Hospital staff, whose collaboration made this case study possible.

medical history, let us look at the psychological test findings in an attempt to discover the psychodynamics that were operating in this case approximately 2 months before her untimely death.

PROCEDURE

The findings to be reported were obtained on the basis of data gathered in a 2-hour, individual, testing session conducted by an experienced clinical psychologist. The patient was administered the following tests: (*a*) vocabulary and picture completion subtests from the Wechsler-Bellevue Intelligence Scale; (*b*) 50-item, true-false, manifest anxiety scale; (*c*) happiness self-rating scales; (*d*) 52-item sentence completion test, containing 26 first-person and 26 third-person stems; (*e*) Cards 1, 2, 3GF, 4, 7GF, 10, 12F, and 13MF from the standard Thematic Apperception Test (TAT); and (*f*) draw-a-person test. The three projective protocols were independently scored for reliability by two experienced clinical psychologists, and a highly satisfactory degree of agreement was found in the scoring and interpretation of this material.

At the completion of the testing session, on the basis of his clinical impressions, the psychologist rated the patient on 12 characteristics contained in a maternal personality rating scale. Each of the ratings was made on a 7-point scale, running from 1 at the socially undesirable end to 7 at the socially more healthy end. The 12 scores were summed, resulting in a total score indicative of a negative or a positive personality picture according to the clinical evaluation.

PSYCHOLOGICAL FINDINGS

On the basis of the intelligence test, it was found that the patient had an IQ of 140, indicating very superior intellect. This finding was supported by the psychologist's impression, based on the clinical interaction, that intellectually she was very bright and functioned at a high level.

On the manifest anxiety scale, the patient received a rather low score of 12. Results obtained from previous testing of normal, healthy, pregnant women have indicated a mean manifest anxiety score of about 16. Studies of randomly selected female, nonpregnant, college undergraduates have also reported mean scores of about 16 on this anxiety inventory. Thus, it appears that the patient was below average in the number of anxiety signs and symptoms she admitted experiencing.

The happiness self-ratings involved comparing oneself with most other people and indicating self-ratings along 4-point scales, running from "much less happy" to "much happier." In response to this procedure, the patient indicated that she regarded herself as "slightly happier" than most people. On the scale that required her to indicate the rating she believed she would receive from friends and family members, the patient again made the rating "slightly happier" than most people. In addition to these general happiness ratings, the patient was asked to make two specific ratings concerned with personal contentment with the state of pregnancy. In response to the question, "How happy or unhappy were you to learn that you were pregnant?" the patient indicated that she was "slightly happy." The second rating was made in response to the question, "Compared to the way you felt before you were pregnant, are you happier or unhappier at present?" and the patient rated herself as "slightly happier." Thus, on the basis of the patient's conscious avowal, it appears that she viewed herself as a fairly happy person, thought that she appeared happy to others, and was content to be pregnant.

Before considering the results obtained from the projective techniques, let us report the clinical psychologist's evaluation of the patient's personality structure according to his "maternal personality ratings." These ratings yielded a total score of 78, which is markedly higher than the mean rating of 50 that was obtained in previous studies of normal, healthy, pregnant women. The attributes on which the patient was rated consisted of such traits as contented, peaceful, cooperative, warm, self-assured, communicative, active, etc. This patient's high score indicates that she made a very favorable impression on the psychologist, and he regarded her as a very well adjusted, emotionally mature, and most adequate individual. In this regard, however, it should be pointed out that previous research has shown highly significant positive correlations between patients' IQ, as measured by objective intelligence tests, and the favorableness of clinical evaluations as indicated by the maternal personality ratings.

Projectively, the patient's psychodynamics appear in a quite different light. The sentence completion test and the TAT were scored for the personality traits of egocentricity, pessimism, distrust, anxiety, and resentment, which previous research has shown to form a personality syndrome termed "alienation." Scoring the patient's sentence completions for indices of these personality dispositions resulted in an alienation syndrome score of 16, which is slightly higher than the mean of 13 obtained in previous analyses of responses obtained from normal pregnant women. More noteworthy, however, is the finding that 12 of the patient's alienation scores were based on completions she gave to sentences that were worded in the third person; only 4 of the alienation scores were in response to first-person sentence stems. It appears, therefore, that on a conscious level the patient presented a quite favorable personality picture, but in response to projective stimuli, she appeared considerably more "alienated" or maladjusted than most pregnant women.

A few examples drawn from her sentence com-

pletion test will serve to evidence the kinds of responses that revealed her conflicts and disturbances. In response to the stem "[illegible] trusts——," she answered "nobody," which suggests a suspiciousness and distrust that were also indicated by other sentence completions. The stem, "Paula feels that the world is——," elicited the completion, "coming to an end," which is probably a sign of extreme pessimism. "I feel anxious when——" stimulated the response, "my sugar is low," and in response to the stimulus, "My greatest fear is——," she responded "getting sick." These projective responses, together with many others that were recorded, indicate that the patient was considerably more anxious, pessimistic, worried about her physical condition, and distrustful than she appeared on the surface.

The TAT revealed even greater evidence of psychopathology. The TAT stories were scored for indices of the five negative dispositions that form the "alienation syndrome." Scoring of this fantasy material resulted in an alienation score of 27, which is considerably higher than the mean alienation score of 18 obtained in previous studies of normal pregnant women. In addition to content, outcomes or story endings were also scored for positive or negative indices. The outcomes of four of this patient's five stories had negative outcomes. With the same weighted scoring system that was employed previously, she received a negative outcome score of 8, which is noticeably higher than our mean of 3 from earlier research. Thus, both the content of her fantasies, stimulated by the TAT pictures, and the nature of the endings she gave to her stories, indicate that this patient was noticeably more alienated, disturbed, depressed, and pessimistic than the majority of pregnant women studied previously.

Some excerpts from her TAT protocol will serve to further support the picture of emotional disturbance and inner turmoil that was gleaned from the projective material. In response to Card 2, the patient told a story about a family in which the husband was dead and the son and mother were left alone. She closed this story with, "The mother looks a little pregnant, but I won't put that in the story. She looks a little old to me." On an unconscious level, she apparently had a very dismal outlook on pregnancy and was trying to suppress any disturbing ideas she may have had about being pregnant. Since the patient was 39 years old, her statement that the pregnant woman looked a little old very likely had personal relevance. It is also noteworthy that this story, in which she tried to avoid coping with fantasies about pregnancy, contained a family in which there had been a death.[2]

The patient began her story to Card 3GF with the statement, "Oh my! This is a sad one." Then she told a most depressing story about a family in which the husband had left the wife alone with a young child. The child had been hurt in an accident, and the mother wanted to call the father but did not know where to locate him. The story ended with, "The doctor has said there is little hope that the boy will survive. It's a very depressing looking picture." Thus, here again, there are indices of family conflict, death, and depression.[3]

In telling a story about the female figure in Card 12F, the patient said that, "Outwardly, she appeared to be very kind and proper, but actually she had deep seated hatreds for most people, and if you could see inside her soul, she would look like the old hag that she really was inside." It seems, then, that this patient was concerned about differences between surface appearance and unacceptable motives operating at deeper levels. The reference to the person's "insides" seems particularly noteworthy in view of her own history of surgery, which undoubtedly caused much more emotional difficulty than appeared on the surface. Moreover, this TAT fantasy provides signs of hostility and resentment that were not evident in the patient's responses to the more objective assessment procedures.

In response to Card 13MF, which usually stimulates sexual fantasies but also is apt to evoke stories of aggression from some individuals, the patient constructed a story in which love, sex, hate, and death were intermingled. The story was about a man who was passionately in love with a very beautiful girl. The girl loved him very much, but he was insanely jealous of her and suspected her of being insincere and untrue to him. In a fit of rage, he killed her. The story was ended on the tragic note that, "He was hit by a great feeling of remorse because he knew he was sorry, that she really did love him; but it was too late then." Again we see an unsatisfying relationship between man and woman, with distrust, extreme hostility,

[2] In looking into the patient's background, a noteworthy anecdote was uncovered. Shortly before the patient was seen for psychological testing, in November, two of her neighbors had died. One was a 38-year-old male who died from ulcers; the other was a 70-year-old woman who died from natural causes. It has been reported that about that time the patient told a close friend that death usually comes in threes, and she believed that she was destined to complete the trio.

[3] After the present report was prepared, it was discovered that this patient's husband had been divorced from his first wife, and, moreover, that he suffered from a manic-depressive condition requiring psychiatric treatment.

death, and depression. These examples from the TAT should suffice to indicate the overwhelming evidence of alienation and psychopathology revealed by the patient's fantasies. The content was consistently morbid and conflict laden, with uniformly desperate outcomes.

The final projective material to be considered here was obtained from the draw-a-person test. The patient drew a picture of a girl with the corners of the mouth turned up in a sort of forced smile and with both eyes tightly closed. A plausible clinical interpretation of this drawing is that the figure was trying to shut out forcefully all awareness of what was transpiring both within and around her and was straining to provide a happy outer appearance. The drawing also contained evidence of considerable anxiety, indicated by a great amount of shading, many irregular lines throughout the figure, and much retracing of lines in the contour of the figure. Thus, both on the basis of objective signs and clinical interpretation, the figure drawing suggests that the patient relied very heavily on defense mechanisms of denial and repression to control her severe inner turmoil and conflict.

From the data obtained from this comprehensive battery of direct and projective assessment techniques, it now seems quite evident that in this patient there was great discrepancy between surface and depth. On the basis of her intelligence test performance, her conscious avowal of manifest anxiety symptoms, happiness self-rating, and verbal interaction with the psychologist, she appeared to be functioning very adequately. However, the projective techniques revealed that beneath this seemingly healthy exterior there was a great store of anxiety, distrust, resentment, and pessimism. In this case, therefore, it is essential to compare the findings derived from both the direct and the projective methods in order to acquire a valid picture of her personality dynamics.

In this regard, one wonders if the picture of emotional maturity, confidence, intelligence, and mature control which she presented, even to an experienced clinical observer, may have served to influence the obstetrician and helped him to accept this physically sick woman as a good risk for pregnancy and normal childbirth. It seems possible that if she had appeared anxious, worried, and doubtful in their interpersonal relationship, in view of her physical condition and the risk involved, the obstetrician might have been more reluctant.

It seems warranted to conclude that this patient was considerably more worried and unhappy than she appeared, and that it was a rather exhausting struggle to maintain her emotional control. Just what having this baby meant to her is not certain. The projective tests provide strong suggestions that her marital relationship may not have been a happy one, and she may have viewed the baby as a means of holding the family together. It is also possible that she wanted to have the baby for her own sake, as a result of some form of counterphobic mechanism, denying fate and reality. Giving birth, at this stage of her existence, would serve to prove that she was still able to function as a woman, that she was not merely existing in the shadow of death but was beginning a new life. At any rate, she wanted to have the baby and fought to minimize the emotional upset generated by the serious risk that was involved.

It may be that this tragic case will make some form of contribution to medical science. Our purpose, as psychologists, however, is to show the sort of contribution that intensive psychological assessment can make in medical research settings. Findings from unique cases such as this may help us toward our goal of ultimately attaining greater scientific understanding of human psychodynamics. From a practical viewpoint, this case vividly emphasizes the need to include both direct (surface) methods and projective (depth) methods in any thorough personality assessment

SUMMARY

This paper presents a psychological case study of a 39-year-old woman who died during pregnancy. Previous to the pregnancy, she had undergone major surgery for stomach cancer and had been strongly advised not to become pregnant. During the fourth month of pregnancy she was administered a comprehensive battery of psychological tests. This testing was done in connection with a National Collaborative Project on perinatal factors in child development, and at the time of testing she was regarded as merely another case in a large sample of pregnant women who were to be included in the research program. During the sixth month of pregnancy, however, the patient died from a brain abscess associated with meningitis, and the baby succumbed following a post-mortem Caesarian delivery. In view of the uniqueness of the case, we conducted a detailed examination of the psychological test findings in the attempt to learn about psychodynamic factors that may have been operating shortly before her death.

The patient's performance on the Wechsler-Bellevue Intelligence Scale showed her to have an IQ of 140. On the manifest anxiety scale, her score indicated fewer anxiety symptoms than the mean number reported in previous groups of normal pregnant women. On a series of self-rating scales,

the patient indicated that she was happier than most people, believed that other people also viewed her as happier than most people, and that she was happy to be pregnant. The examining psychologist's ratings of the patient, on a series of maternal personality characteristics, resulted in a score markedly higher than the average score obtained in a previously studied sample of pregnant women, indicating a very positive personality picture according to the clinical evaluation.

On the contrary, however, findings derived from the TAT, sentence completion test, and draw-a-person test, provided abundant evidence of psychopathology underlying the surface picture of healthy emotional adjustment. It was found that the patient was extremely anxious, distrustful, resentful, and pessimistic. Thus, the projective material revealed that she was an alienated and conflict-laden individual who made strong use of the mechanisms of denial and repression in the attempt to hide her inner turmoil both from her own conscious awareness and from the awareness of others. It was concluded that these case findings serve to reaffirm the importance of including both direct methods and projective methods in any thorough assessment of personality dynamics.

(Early publication received March 21, 1960)

Journal of Abnormal and Social Psychology
1960, Vol. 61, No. 3, 292-294

CRITIQUE AND NOTES

THE COMPLEMENTARITY OF PERSONALITY NEEDS IN FRIENDSHIP CHOICE

MARY ST. ANNE REILLY, WILLIAM D. COMMINS, AND EDWARD C. STEFIC
Catholic University of America

EARLY studies in the dynamics of friendship agreed that friendship was characterized by a positive, though slight, resemblance between friends. However, the personality theories of Murray (1938) and Rogers (1951) suggested that friendship involves a complementary relationship of mutual need satisfaction, and Winch, Ktsanes, and Ktsanes (1955) found evidence of complementary needs in mate selection. The present study, concerned with investigating the complementarity of personality needs in friends of the same sex, proposed these questions: Are the need patterns of friends complementary? Are the personality needs of friends mutually satisfying?

The two main hypotheses proposed by Winch et al. (1955) served as a framework. These were that (*a*) a person high in one need would be attracted to someone low in that need (complementarity of needs), and (*b*) a person high in one need would be attracted to someone high in the opposite need (need satisfaction). In addition to investigating the actual, self-perceived need patterns of friends, we were interested also in studying the phenomenal need patterns of friends—their needs as perceived by each other. Here, we were following Rogers' (1951) suggestion: "it is the perception, not the reality, which is crucial in determining behavior." Finally, to establish a basis for comparison with earlier studies in the field, of which some have found friends to be similar in values or evaluative attitudes, we proposed a third hypothesis, that friends would tend to be similar in values.

METHOD

Two groups of resident female Catholic college students volunteered as mutual friend pairs, with 25 junior pairs and 25 sophomore pairs (total $N = 100$). A control group, set up by randomly matching the juniors with the sophomores by a table of random numbers, formed a contrasting group of nonfriends. The age range of the subjects (*Ss*) was from 18 years 11 months to 21 years 1 month, with a mean age of 19 years 11 months.

The first instrument used was the Edwards Personal Preference Schedule (PPS), which provides measures of 15 personality needs. The PPS was designed to minimize the factor of social desirability, by employing a forced-choice format. The second instrument was the Allport-Vernon Study of Values (1931), which provides relative measures of six basic interests or values. (The 1931 edition, instead of the revised edition was used, as a better basis of comparison with earlier studies in the area.)

The PPS was administered to all junior *Ss*, to obtain self-perceived need scores (each *S* took the test as she perceived herself). During the same testing period, the Allport-Vernon was also administered. Two days later, the same *Ss* took the PPS, but this time as each thought her friend would answer. This yielded friend-predicted need scores for all *Ss*. Two weeks later, all sophomore *Ss* took the tests, following the same procedure.

From the raw scores of the personality needs measured by the PPS, Pearson product-moment correlations were computed for the following two variables of friends: (*a*) the self-perceived needs of friends, and (*b*) the phenomenal needs of friends (an *S*'s self-perceived needs with that *S*'s prediction of her friend's needs); and one variable on nonfriends: their self-perceived needs. From the raw scores on the six Allport-Vernon values, correlations were computed for an additional variable on friends: self-perceived values.

RESULTS AND DISCUSSION

The first hypothesis, about complementary needs between friends, was tested via the following questions:

Does a person have for a friend, one whose actual, self-perceived needs complement her own?

Is there any difference between the correlation of a friend pair's actual self-perceived needs, and that friend pair's phenomenal needs (the correlation of a person's actual needs with her prediction of her friend's needs)?

What is the relationship of the personality needs of nonfriends?

The following correlations on friends and nonfriends were inspected: complete need profiles, each personality need with the same need in the partner, and the highest need of each *S* with the same need in the partner. A complementary relationship between the personality needs should be indicated by negative correlations.

Need Profiles

In respect to the 50 correlations of the actual, self-perceived need profiles of the 50 friend pairs, and those of the 50 nonfriend pairs, the number of negative *r*'s was the same for both groups. The range of the *r*'s for friends was from $-.54$ to .85

with a mean r of .23, while the range for nonfriends was from −.54 to .85 with a mean r of .18, the difference obviously falling short of significance. This finding was not similar to that of Winch et al. (1955), who reported a significant difference between the mean r of .10 on husband-wife pairs in a need interview, and .25 for randomly matched men and women. Admittedly, the husband-wife relationship may be more of a complementary one than that between friends.

An inspection of the 100 correlations of friend pairs' phenomenal need profiles shows that the proportion of negative correlations here was almost the same as for the self-perceived need profiles of friends; the range of r's was from −.73 to .90 with a mean r of .25. This value of course does not differ significantly from that of .23 for the self-perceived need profiles of friends.

From the data, then, it is clear that the self-perceived need profiles of friends were not complementary to one another. There was no significant difference between friend pairs' actual needs and the phenomenal needs of friend pairs. Need profiles of nonfriends did not differ significantly from those of friends.

If these data in regard to need profiles did not support the first hypothesis of a complementary relationship between friends, neither did they warrant any assertion as to the similarity of friends in respect to personality needs. In the first place, though the mean r of the correlations of friends' need profiles was positive (.23) it was not significantly different from that of nonfriends (.18).

TABLE 1

CORRELATIONS OF FIFTEEN PERSONALITY NEEDS

Personality Needs	Friends' Actual	Friends' Phenomenal	Nonfriends' Actual
Achievement	.09	.24*	−.07
Deference	.16	.25*	−.13
Order	−.05	−.03	.01
Exhibition	.06	.22*	−.14
Autonomy	.11	.31**	.13
Affiliation	.16	.20*	.18
Intraception	−.07	.14	.04
Succorance	.16	.19*	−.10
Dominance	.17	.15	.11
Abasement	.13	.35**	.05
Nurturance	.02	.30**	−.02
Change	−.15	.10	.12
Endurance	−.15	−.09	.15
Heterosexuality	−.17	.32**	.09
Aggression	.08	.29**	−.14

* Significant at the .05 level.
** Significant at the .005 level.

TABLE 2

CORRELATIONS OF OPPOSITE NEEDS OF FRIEND PAIRS

Opposite Needs	Actual	Phenomenal	Opposite Needs	Actual	Phenomenal
Ach.-Def.	.02	−.12	Aut.-Suc.	.01	−.09
Ach.-Aff.	.02	.15	Aut.-Aba.	−.07	−.09
Ach.-Suc.	.00	.10	Aut.-Nur.	−.01	−.24
Ach.-Aba.	.07	−.11	Aff.-Dom.	.05	−.05
Ach.-Nur.	−.01	.03	Aff.-Agg.	.11	−.12
Def.-Exh.	−.17	.07	Dom.-Suc.	−.15	−.19
Def.-Aut.	−.13	−.03	Dom.-Aba.	−.09	.16
Def.-Dom.	.08	.00	Dom.-Nur.	.02	.52
Def.-Agg.	.01	−.12	Agg.-Aba.	.02	−.15
Aut.-Aff.	.06	−.29	Agg.-Nur.	.12	−.08

Separate Personality Needs

The second step in the analysis sought to discover whether friends tend to be more complementary to one another in regard to certain personality needs than to others. To this end, correlations for each of the 15 personality needs from the data on the self-perceived needs of friends and of nonfriends, as well as from the phenomenal needs of friends, were inspected (see Table 1). The double entry method of computing product-moment correlations for each of the 15 personality needs was used, as recommended by Goodenough and Anderson (1931), for measuring an interchangeable variable. No consistent complementary relationship between friends on personality needs was indicated. In other words, though some friend pairs were complementary on need endurance, for example, other pairs were similar in regard to this need. There was no significant difference between the r's of friends and nonfriends; or between the r's of friends' self-perceived needs and their phenomenal needs.

Each S's Highest Need

The data were next analyzed to find out whether the need on which the *S* obtained the highest score was complementary to the same need in the partner. *S*'s highest need correlated .02 with the same need from the friend's self-perceived needs, .01 with the prediction of her friend's need, and −.01 with the nonfriend's self-perceived need, values not significantly different from one another or from zero. There was no evidence, therefore, of a complementary relationship between a person's highest need and that same need in her friend.

To test the second hypothesis about mutual need satisfaction between friends, opposite needs of friend pairs were compared, our prediction being that they would correlate positively. For this problem, a list of 20 sets of opposite needs

TABLE 3
VALUES OF FRIEND PAIRS

Values	*r*	*M*	*SD*
Theoretical	.01	24.8	6.18
Economic	.09	23.4	6.92
Aesthetic	.14	27.3	7.16
Social	.06	29.4	5.01
Political	.15	30.5	5.92
Religious	.03	42.6	6.82

was devised from statements from Murray (1938, pp. 151–181). (Edwards has reported in the table of intercorrelations in his PPS manual, all negative *r*'s, ranging from −.12 to −.36, on the opposite needs which were used in this study. However, these *r*'s are not comparable to those presented in the present study, since Edwards' *r*'s are intra-individual, while the *r*'s obtained in this sample are between friends.) The data were analyzed in regard to the self-perceived needs of friends, and the phenomenal needs of friends, as is indicated in Table 2. Though a number of positive *r*'s were obtained, these were too low to support the hypothesis of mutual need satisfaction between friends. Here again, this finding was not what might have been expected as a result of the findings of the somewhat similar study on mate selection of Winch et al. (1955). They reported, for example, that "men high in need achievement tend to select as wives, women who are high in need abasement, need deference, need succorance, and need nurturance." We suggest that while the essential nature of the heterosexual relationship of husbands and wives may be a complementary one, that of friends of like sex is not essentially complementary.

In this sample, data on the values of friends from the Allport-Vernon Study of Values were very similar, on the whole, to earlier studies in this area. In regard to Richardson's study (1940), for example, the *r*'s were similar, except for theoretical and religious values, for which he found *r*'s of .21 and .34, respectively. However, because the present sample was more homogeneous, restriction of range could account for the lower correlations. The means of each of the values were also similar in the two studies, the only exception being that for religious value, which was 42.6, significantly higher than the 31.4 reported by Richardson. The data indicate a positive though slight resemblance between friends in values, as can be seen in Table 3.

SUMMARY AND CONCLUSIONS

This study sought to find out whether friendship involves a complementarity of personality needs, and whether these needs are mutually satisfying. Need profiles from the Edwards PPS were obtained for each of 50 friend pairs. Friend-predicted need profiles from the PPS were also obtained. Relative scores of six values from the Allport-Vernon Study of Values were secured for each *S*. Correlations were computed for complete need profiles, for single personality needs, for opposite needs, for the three following groups: friends' self-perceived scores, friends' self-perceived with friend-predicted scores, and randomly matched pairs' self-perceived scores.

The following conclusions were drawn:

1. In this sample, no consistent complementary relationship was found in regard to self-perceived personality needs of friends, nor was there any evidence of mutual need satisfaction between friends.

2. Friends, in general, did not tend to see themselves and their friends as more consistently complementary rather than similar.

3. On the other hand, there was no conclusive evidence for a relationship of similarity of personality needs of friends.

4. Friends tended to be slightly similar in values.

REFERENCES

GOODENOUGH, FLORENCE L., & ANDERSON, J. E. *Experimental child study*. New York: Century, 1931.

MURRAY, H. A. *Explorations in personality*. New York: Oxford Univer. Press, 1938.

RICHARDSON, H. M. Community of values as a factor in friendship of college and adult women. *J. soc. Psychol.*, 1940, **11**, 303–312.

ROGERS, C. R. *Client-centered therapy*. New York: Houghton Mifflin, 1951.

WINCH, R. F., KTSANES, T., & KTSANES, VIRGINIA. Empirical elaboration of the theory of complementary needs in mate selection. *J. abnorm. soc. Psychol.*, 1955, **51**, 508–513.

(Received June 22, 1959)

Journal of Abnormal and Social Psychology
1960, Vol. 61, No. 2, 295-297

CONTEXT EFFECTS IN SOCIAL PERCEPTION[1]

LEON H. LEVY
Indiana University

BEGINNING with the now classical study by Murray (1933) on the effects of fear arousal on judgments of maliciousness in photographs, a fairly extensive body of literature has developed concerned with the personal determinants of social perception. In contrast, little is known concerning the situational determinants of social perception, especially those which do not seem to have a readily apparent effect upon one of the presumed personal determinants. That is, excluding situational factors that might be expected to induce some affective response in the perceiver such as fear, frustration, or hostility, and in contrast to our knowledge of the effects of various field factors such as background illumination and ambient noise level upon perception of relatively simple visual and auditory stimuli, our knowledge of the pertinent situational determinants of social perception is grossly deficient. Yet there must be such determinants and any theory of social perception cannot be considered complete without taking them into account.

Hunt (1941), Helson (1948), and others (Rudin & Stagner, 1958) have argued that laws of judgment exist which are not unique to the type of stimuli judged. Hunt (1941), attempting to establish this point, demonstrated shifts, allegedly reflecting anchor effects, in judgments using an 11-point rating scale in which the ends were either anchored or not depending upon the experimenter's (*E*'s) instructions to the subject (*S*) regarding their meaning. Unfortunately, the results obtained by this procedure would seem to be just as easily accounted for by alterations in verbal response probabilities brought about by the instructions, as by alterations in perceptual processes.

The present study continues within the vein of that by Hunt (1941) but differs from his in that it studied anchor effects where these were effected entirely through modifications of the stimulus context without alterations in verbal instructions. Finding predictable judgmental shifts with this type of procedure would seem to present an especially strong argument for the presence of field determinants of social perception as well as for the nonspecificity of the laws of judgment. This approach clearly places the locus of effect within the stimulus situation, obviates the problem of dealing conceptually with verbal instruction, and reduces the applicability of an interpretation in terms of response probability.

The hypothesis to be tested in this study is that judgments of a target photograph vary inversely with the context of other photographs with which it is presented, such that the greater the probability that the context will elicit judgments in one direction on a particular bipolar dimension, the greater the probability that the target will be judged in the opposite direction. This hypothesis is predicated upon the assumptions that (*a*) judgments of this type are a function of the location of a point of subjective equality or adaptation level (Helson, 1948) within any given dimension and (*b*) that this point varies directly with the value of the stimuli comprising the judgmental context.

The hypothesis was tested by using four bipolar dimensions, good–bad, dangerous–safe, reliable–unreliable, and tense–relaxed, and three contexts for each designated as high, medium, or low in terms of the estimated probability that the stimuli comprising the context would be judged as good, dangerous, reliable, or tense, respectively.

METHOD

Stimuli. The photographs used in this study were of white, male applicants for admission to graduate training in psychology at Indiana University who were either not accepted or did not elect to come. Each of these pictures had been used in a previous study (Levy & Dugan, 1960) in which 100 *S*s had judged them in terms of the traits used in this study as well as other traits. Since, in this previous study, these pictures were part of a total of 225 presented in random order and judged on one of 15 bipolar dimensions, it did not seem likely that context effects, as they were to be studied here, entered to any great extent in the judgments obtained. Thus records of the frequency of judgments made for each picture in the previous study served as the basis for selecting pictures that comprised the various contexts and targets in the present study.

For each dimension, seven pictures were used, two for each context and one as the target, with no picture being used in connection with more than one dimension. Each picture was mounted on a 3 × 4 inch piece of poster board with a label containing the two terms of the dimension attached below it. Judgments of photographs constituting the high context condition had a frequency range between 80% and 98% in the specified direction for the dimension in question, for medium con-

[1] This investigation was supported in part by a Research Grant (M-2067) from the National Institute of Mental Health of the National Institutes of Health, United States Public Health Service.

The author wishes to express his gratitude to Jane A. Kelly who ran the subjects of this study.

text pictures the corresponding judgments ranged between 46% and 56%, and for low context pictures between 10% and 35%. Target photographs ranged between 49% and 52% in judgments placing them at one particular end of their respective dimensions.

Subjects. The *Ss* used in this experiment were 120 students, 78 men and 42 women, drawn from an introductory psychology course. They were assigned at random to one of four groups of 30 *Ss* each, with the proportion of men and women approximately the same between groups. None of these students had participated in the previous study using the photographs.

Procedure. Each *S* was seen individually and presented with three sets of three pictures each, one set at a time. All pictures within a set were judged on the same dimension and consisted of two context pictures and one target. Each set represented a different dimension as well as a different context level, and within each group the order of presentation of context levels, and consequently dimensions, was systematically counterbalanced. Thus all *Ss* in any given group made judgments in terms of only three of the four dimensions and in connection with only one corresponding context for each, but over all four groups, each dimension was involved at each context level equally often yielding 30 judgments for each target at each context level of its respective dimension.

On each presentation, the set of three pictures was presented on a small stand directly in front of the *S* with the target picture flanked by each of the context pictures. *Ss* were instructed to look at all three pictures and then to give their first impressions of each in terms of one of the two adjectives on the label below the picture. As soon as this was done and recorded by *E*, the next set was presented. From the *Ss'* standpoint, the target pictures were in no spurious way distinguishable from their contexts.

RESULTS

The number of *Ss* judging each target as good, dangerous, reliable, or tense, was tabulated for each context level. These data, presented in Table 1, reveal the monotonic increase predicted from the hypothesis in all instances except the good–bad dimension where a reversal between the high and medium context levels occurred. A Friedman two-way analysis of variance by ranks (Siegel, 1956) applied to these data in order to determine whether judgments differed significantly between context levels over all four dimensions, yielded a χ^2 of 6.5 which is significant at the .04 level. Further, testing for the significance of the difference in proportions of judgments in a specified direction between the high and low context levels for each dimension separately indicated a difference significant at the .04 level for the dangerous-safe dimension and at the .003 level for the tense-relaxed dimension, using a one-tailed criterion in each case since the direction had been predicted. Thus, these results clearly support the hypothesized inverse relationship between the probability of particular judgments of the context photographs and of targets contained in the context.

TABLE 1
FREQUENCIES OF SUBJECTS MAKING SPECIFIED JUDGMENTS OF TARGET PHOTOGRAPHS IN VARIOUS CONTEXTS

Dimension	Context Levels			Significance of proportion of high vs. low judgments
	High	Medium	Low	
Good–Bad	13	12	16	*ns*
Dangerous–Safe	14	17	20	.04
Reliable–Unreliable	12	15	18	*ns*
Tense–Relaxed	8	16	18	.003

Note.—The frequencies given are for *Ss* giving the term in italics as their judgment. In each case the total possible judgments is 30.

DISCUSSION

In the language of social psychology, the present findings have most immediate relevance for the concept of reference groups (Hyman, 1942; Kelley, 1952; Merton & Kitt, 1950; Newcomb, 1950; Sherif, 1948). For the *Ss* of this experiment, the context pictures may be seen as functioning as reference groups for their judgments of the target pictures. From this standpoint, the results clearly point to the potency of the reference group as a standard of comparison in social perceptual situations even where the likelihood is minimal that the person will encounter any sanctions for ignoring it. Further, membership in a group is clearly not a necessary condition for its functioning as a standard of comparison for the individual. Indeed, the principles governing the operation of reference groups in social perception appear to be no different from those operating in classical psychophysical situations involving judgments of standard and comparison stimuli.

Certain questions that might be raised in connection with the operation of reference groups would seem to be ideally suited to investigation by the method used in the present study. By appropriate manipulations of the constitution of the set of context pictures, the effects on judgments of such variables as the size of the context, context-to-target social relationships of various kinds, heterogeneity of contexts, etc. are open to investigation. The conditions under which context effects are maximal and minimal may also be studied by this method.

Implicit in the latter problem is the question of individual differences in susceptibility to context effects and their personality correlates. The extent to which the context or reference group affects a person's judgment would seem akin to the question of field dependency in personality theory (Witkin,

Lewis, Hertzman, Machover, Meissner, & Wapner, 1954), and thus an opportunity exists, by means of the present technique, to test the generality of this construct and also to effect a further *rapprochement* between social psychology and personality theory.

Personality theories in which the self-concept plays a central role may also find the results of the present study of some relevance. For if context affects judgments of others, it is reasonable to believe that it would have a similar effect upon self-perception. Efforts to account for a person's self-concept at any given time, as well as attempts at altering it, might therefore do well to consider the context within which the person judges himself as well as his personality dynamics and life history.

SUMMARY

Concerned with the problem of field determinants of social perception, this study investigated contextual effects on judgments of photographs of persons. Photographs, designated as targets, were presented in contexts of two other photographs which were expected, on the basis of previous research, to vary in their probability of eliciting a particular judgment along a specified bipolar trait dimension. Data obtained clearly supported the hypothesized inverse relationship between the probability of the context eliciting a particular judgment and that of the target eliciting a similar judgment. The results were discussed in connection with reference group theory in social psychology and the constructs of field dependency and the self-concept in personality theory.

REFERENCES

HELSON, H. Adaptation-level as a basis for a quantitative theory of frames of reference. *Psychol. Rev.*, 1948, **55**, 297–313.

HUNT, W. A. Anchoring effects in judgment. *Amer. J. Psychol.*, 1941, **54**, 395–403.

HYMAN, H. H. The psychology of status. *Arch. Psychol., NY*, 1942, No. 269.

KELLEY, H. H. Two functions of reference groups. In G. E. Swanson, T. M. Newcomb, & E. L. Hartley (Eds.), *Readings in social psychology*. (Rev. ed.) New York: Holt, 1952. Pp. 410–414.

LEVY, L. H., & DUGAN, R. D. A constant error approach to the study of dimensions of social perception. *J. abnorm. soc. Psychol.*, 1960, **61**, 21–24.

MERTON, R. K., & KITT, ALICE S. Contributions to the theory of reference group behavior. In R. K. Merton & P. F. Lazarsfeld (Eds.), *Studies in the scope and method of "the American soldier."* Glencoe, Illinois: Free Press, 1950. Pp. 40–105.

MURRAY, H. A. The effect of fear upon estimates of the maliciousness of other personalities. *J. soc. Psychol.*, 1933, **4**, 310–329.

NEWCOMB, T. M. *Social psychology*. New York: Dryden, 1950.

RUDIN, S. A., & STAGNER, R. Figure-ground phenomena in the perception of physical and social stimuli. *J. Psychol.*, 1958, **45**, 213–225.

SHERIF, M. *An outline of social psychology*. New York: Harper, 1948.

SIEGEL, S. *Nonparametric statistics*. New York: McGraw-Hill, 1956.

WITKIN, H. A., LEWIS, H. B., HERTZMAN, M., MACHOVER, KAREN, MEISSNER, P. B., & WAPNER, S. *Personality through perception*. New York: Harper, 1954.

(Received July 2, 1959)

Journal of Abnormal and Social Psychology
1960, Vol. 61, No. 2, 297-300

THE FORMATION OF POSITIVE ATTITUDES TOWARD GROUP MEMBERS[1]

BERNICE EISMAN LOTT AND ALBERT J. LOTT
Kentucky State College *University of Kentucky*

IN an attempt to predict group behavior on the basis of general psychological principles as contrasted with formulations specific to the area, the concept of group cohesiveness has been re-examined and reformulated within a learning theory framework (Lott, in press). Within this framework, cohesiveness is defined as *that group property which is inferred from the number and strength of mutual positive attitudes among the members of a group*. The concept of attitude is used, instead of the more usual one of attraction, because of its precise and particular meaning within learning theory, i.e., an implicit anticipatory response having cue and drive properties (Doob, 1947). By defining cohesiveness in terms of mutual positive attitudes among group members, the assumption is made that the members who comprise a group constitute its most significant components.

A number of hypotheses regarding both the antecedents and consequents of cohesiveness, as defined above, have been derived from learning theory (Lott, in press). The problem of this experiment has been to test the most fundamental

[1] A paper, based on this investigation, was read at the 1959 American Psychological Association meetings in Cincinnati.

of these hypotheses, one which concerns the conditions under which positive attitudes toward group members may be formed.

It is predicted that if a person is rewarded in the presence of others (fellow group members), he will develop positive attitudes toward them. This proposition rests upon the following assumptions:

1. Persons may be conceptualized as discriminable stimuli to which responses may be learned.

2. A person who experiences reinforcement or reward for some behavior will react to the reward, i.e., will perform some observable or covert goal response (R_g or r_g).

3. This response to reward will become conditioned, like any other response, to all discriminable stimuli present at the time of reinforcement.

4. A person (group member) who is present at the time that Individual X, for example, is rewarded thus becomes able, in a later situation, to evoke R_g or, what is more likely, its fractional and anticipatory component, $r_g - s_g$. This latter response, which Hull has called "expectative" (1952, Ch. 5), was earlier interpreted by Doob (1947) as the underlying mechanism of an attitude.

The specific prediction tested by this study is that members of three-person groups who are rewarded for their performance in the presence of their fellow group members will more likely develop positive attitudes toward them than will members of such groups who are not rewarded. Positive attitudes are inferred, here, from choices made on a sociometric test subsequent to, and outside of, the experimental situation.

METHOD

Subjects

Forty-eight children from the University of Kentucky Elementary School, 24 each from Grades 3 and 5, served as *Ss*.[2]

Procedure

The *Ss* were divided into 16 three-member groups, following the administration of two sociometric tests. These tests were given, by the regular classroom teachers, on two consecutive days, several days before the actual experimental situation. On the basis of the test results the groups were formed so that each group was made up of children who had *not chosen each other* on either of the tests. Four all male and four all female groups were formed from Grade 3; five male and three female groups were formed from Grade 5.

Omitting the preliminary instructions, the criterion question asked in the first test was as follows:

> *Test I.* Let us suppose that each one of you gets picked to take a trip to the moon in a rocket ship. This is a very special trip .... It is important for everyone who is on the same ship to get along well with each other .... Because of this you get a chance to pick two children to go along with you. Now, of all the children in this class, which two would you pick to travel in the same rocket ship with you?
>
> *Test II.* [For this test the children were asked to think of themselves as visitors from Mars and to choose two classmates whom they would like to have waiting for them when their spaceship landed on Earth.]

For a group situation in which rewards and nonrewards could be manipulated, a board game called "Rocket Ship" was devised. The object of the game, played by groups of three, is to land cardboard rocket ships on planetary objectives. Each objective is reached by traversing a separate path containing four danger zones, at each of which a choice between a white and a striped subpath (one "safe" and the other "dangerous") must be made. By having the children in a group take turns crossing the danger zones first, the *Es* could arrange to have some children succeed and others fail in reaching the planets safely. The manner in which this was accomplished will be described below.

On the day of the experiment proper (separate days for the two classes), the *Es* were introduced to *Ss* as having developed a children's game which they wished to test. One group at a time was called upon (in a predetermined random order) to play the game in a room adjoining the regular classroom.[3] The following instructions, given to each group by one of the investigators (E_1), describe the manner of playing:

> ... You three are rocket ship pilots. Each one of you has your own ship but you are going to take a trip into outer space together, side by side. The first trip you are going to take is to ——. In order to get there you must follow this path which scientists have decided is the safest way to go. The scientists also know that, at a few points along the way on this path, there is great danger. When you get to one of these dangerous points you are going to have to stop and decide whether the striped path or the white path is the dangerous one. One of the paths will get you past the danger safely. If you take the wrong path your rocket ship will be blown up and you'll have to parachute to Earth.... When you get to a danger point you'll have to decide which one of the group will take a chance and be the first to try either the striped or white path. If he gets through safely, the other two rockets may follow him. If his ship should get blown up, though, the other two ships will take the other path which you will know to be safe. You'll know that a ship has been blown up when

[2] The authors wish to acknowledge the cooperation of the following persons whose kind assistance made possible both the reported experiment and the pilot work which preceded it: M. Hitch, Principal of Rosenwald Laboratory School, Frankfort, Kentucky; A. Wolfe, intermediate grade teacher at Rosenwald; E. Sasman, Principal of University Elementary School, Lexington, Kentucky; A. Boone, M. Moore, and O. Barrett, teachers at University School of the third, fourth, and fifth grades, respectively.

[3] Wherever a child, preselected for a particular group, was absent on the day of the experiment, a substitute who met the criterion of nonchoice by, and of, the other group members, was, in all cases, available.

you hear the sound of a bell, like this [sounded by E_2]. . . . Remember, you must take turns being first. . . . It's possible for all three of you to get through to a planet safely, but it may be that only two of you will make it, and maybe just one or none of you will make it. . . . If you reach a planet safely you will be able to choose one of these prizes [small plastic auto models] which you may keep. . . . We'll play half the game this morning and half this afternoon.

Each group tried for three objectives in the morning and three in the afternoon. Half the total number of *S*s was permitted to land safely on four planets (two at each session) while half was prevented from reaching any of them. "Reward" was thus defined as the receipt of four plastic car models (one for each successful landing). A child was either "rewarded" or "not-rewarded," i.e., made no successful landings at all and received no prizes; there were no in-between conditions. Prizes won in the morning session were held by the *E*s until the end of the afternoon session at which time the rewarded *S*s returned to their classroom with four model cars and the nonrewarded *S*s returned with none.

Rewarded and nonrewarded *S*s were selected on a random basis prior to the game. A nonrewarded *S* always had his ship "blown up" when, at his chance to go first at a danger point, he took either the striped or white path.[4] As soon as he made his choice of path, the "blow up" bell was rung by E_2 who sat somewhat apart from the game area.

Each group had been randomly assigned, prior to the game situation, to one of the following conditions which describe the number of group members who were to be rewarded during the game: zero, one, two, all. Four groups, two from each class, were assigned to each condition. This aspect of the design was introduced to avoid having the *S*s suspect that the game was "rigged." The natural flavor of the game was maintained by having *S*s know that it was possible for all three group members to be successful, or only two, etc.

Shortly before the close of the school day, approximately one hour after the last group had played the game, the classroom teacher administered another sociometric test (III), as follows:

> Suppose your family suddenly got the chance to spend your next vacation on a nearby star out in space . . . you can invite two children to go on the trip . . . and spend the holiday with you on the star. Which two children in this class would you choose to take with you?

After the choices were collected by the teacher, E_2 appeared before the class to thank the *S*s for their cooperation and help. And, because everyone had been such "good sports," four prizes were distributed to each of the youngsters who had not won them during the game.

[4] Each child was in the position of "being first" at a danger point at least once on his way to each of the planets since *S*s had been instructed to take turns and there were four danger zones per path and three children in a group. It was always possible, therefore, to "blow up" nonrewarded *S*s without interfering with a group's spontaneous behavior regarding which child would go first at which point.

TABLE 1
CHOICES MADE BY SUBJECTS ON SOCIOMETRIC TEST III

Subjects	Choices		N^a
	Play-group member	Non-play-group member	
Rewarded	11	37	48
Nonrewarded	3	45	48

[a] N = number of choices made; each *S* made two choices.

RESULTS

The results of the final sociometric test (III), which succeeded the play-group experience, are presented in Table 1. The proportion of play-group members chosen by rewarded *S*s was found to be significantly greater than the proportion chosen by nonrewarded *S*s. The obtained critical ratio, corrected for continuity, is 2.14 ($p = .03$; two-tailed). This confirms the prediction that *S*s who had been rewarded would choose members of their groups, on the final sociometric test, significantly more often than *S*s who had not been rewarded.

DISCUSSION

The present findings indicate that the formation of positive attitudes toward persons is predictable from learning theory principles and can be studied in the laboratory. This study thus extends the applicability of a general S-R framework to significant social behavior.

The prediction that positive attitudes toward persons can be formed by experiencing reward in their presence, was clearly confirmed. Since it is in terms of such positive attitudes among the members of a group that cohesiveness has been defined, the present experiment is seen to be concerned with the antecedents or determinants of cohesiveness even though no attempt was made to measure the variable directly. The specific concern of this first study was with the development of positive attitudes toward group members and not with the group property that results from such attitudes when they are mutual.[5]

There have been comparatively few experimental attempts to vary group cohesiveness other than by suggesting to the members of a group that they will like each other. One investigation in which determinants were experimentally manipulated (Thibaut, 1950) found a positive relationship between cohesiveness and group status. In another study, more relevant to the present one

[5] An operational measure of cohesiveness which follows directly from the definition of the concept given earlier has been developed by the investigators and will be reported in connection with its use in another study.

(Bovard, 1951), a significantly greater level of interpersonal affect was found in group-centered as compared with traditional leader-centered classes. Bovard suggested that this result was due to the fostering, in group-centered classes, of member-to-member interaction which produces greater accuracy in intermember perception, creating, according to Bovard, a situation conducive to "need satisfaction." That the experience of need satisfaction (reward) in the presence of group members can, indeed, result in positive affect toward those group members has been demonstrated in the present experiment.

The results obtained in this study should be evaluated in the light of the following factors which could only have tended to work against substantiation of the hypothesis:

1. Play-group members were, by design, neutral or negative stimuli for one another at the beginning of the experimental situation. This was assured by placing, in a group, only *S*s who had not chosen one another on two previous sociometric tests.

2. The amount of reward experience received by rewarded *S*s in the presence of others was extremely small when compared with the amount of daily contact our *S*s typically had with each other in the classroom, playground, and after school. (Practically all the *S*s in each of the classes had been together for their entire schooling, beginning with kindergarten.)

3. The choices made on the two pre-experimental sociometric tests (I and II) were found to be unexpectedly stable for third and fifth graders, indicating relatively reliable friendship ties[6] which the experimental experience was able to disrupt, however temporarily.[7]

[6] Forty percent of the *S*s made the same two choices on both tests; another 42% of the *S*s repeated one choice on both tests.

[7] No follow-up test was given since there was no reason to expect that the positive attitudes formed during the brief game situation would last, in the absence of continued reward experience in the presence of the same individuals.

Despite the above factors the prediction, that positive attitudes toward persons will be developed as a result of the receipt of reward in their presence, was clearly supported. Though the results were obtained with children, there is no reason to expect that they would not hold with adults as well. Such generalization must, of course, await adequate test.

The variable manipulated in the present study was simply reward vs. nonreward. Future research might profitably deal with variations in reward frequency, delay, and schedule, for example, as these affect the conditioning of attitudinal responses, and, consequently, as these affect the development of cohesiveness in small groups. That the results obtained in this study were predicted from general behavior principles increases our confidence in the promise which this kind of an approach holds for the general area of small group behavior.

SUMMARY

Three-member groups of children played a game in which some members were rewarded and others were not. On a later sociometric test, outside of the game situation, rewarded *S*s chose a significantly greater proportion of their fellow group members than did the nonrewarded *S*s. These results were predicted from general principles of S-R learning theory.

REFERENCES

BOVARD, E. W. The experimental production of interpersonal affect. *J. abnorm. soc. Psychol.*, 1951, **46**, 521–528.

DOOB, L. W. The behavior of attitudes. *Psychol. Rev.*, 1947, **54**, 135–156.

HULL, C. L. *A behavior system*. New Haven: Yale Univer. Press, 1952.

LOTT, BERNICE E. Group cohesiveness: A learning phenomenon. *J. soc. Psychol.*, in press.

THIBAUT, J. An experimental study of the cohesiveness of underprivileged groups. *Hum. Relat.*, 1950, **3**, 251–278.

(Received July 20, 1959)

PSYCHOLOGY AND REHABILITATION

Beatrice A. Wright, Ph.D.
Editor

Based on the proceedings of an Institute on the roles of psychology and psychologists in rehabilitation, sponsored by the American Psychological Association and supported by a grant from the Office of Vocational Rehabilitation, Department of Health, Education and Welfare.

This volume examines the principles and assumptions underlying the modern philosophy of rehabilitation, paying special attention to the roles and functions of psychologists as members of an interdisciplinary team. Problems of training are explored in detail and in the context of the changing nature of the field. Areas of needed research are highlighted.

Price $1.50

Order from

American Psychological Association, Inc.

Dept. PAP
1333 Sixteenth Street, N.W.
Washington 6, D. C.

WILEY

A new, important book in which 30 outstanding specialists in the field of child development offer the first guide to research methods . . .

HANDBOOK OF RESEARCH METHODS IN CHILD DEVELOPMENT

Edited by PAUL HENRY MUSSEN, *Associate Professor of Psychology, University of California, Berkeley.* This is the only book in Child Development that deals with broad, general issues of research as well as with specific research techniques. Part I is concerned with general research design and use of observation and experimental methods in research. The next section examines psychological growth and development. Part III is devoted to the study of cognitive processes (perception, learning, language, etc.). In Part IV the contributors deal with the major research tools and areas of study in the fields of personality development and social behavior. The final section studies the child's social environment.

In a sense, *Handbook of Research Methods in Child Development* is addressed directly to present and potential investigators who can use it as a basic reference and a guide to research matters outside their own research specialties. Also, it is valuable to advanced scholars and graduate students interested in a balanced picture of the techniques applied in child study. Subjects treated are of broad interest and will be useful as source material in courses devoted to research methodology in psychology and other behavioral sciences.

Contributors include:

Anne Anastasi	Bärbel Inhelder	Vivian Olum
Donald M. Baer	Orvis C. Irwin	Austin W. Reisen
Alfred J. Baldwin	Harriet J. Kelly	Henry N. Ricciuti
Jean Berko	William Kessen	Charles C. Spiker
Sidney W. Bijou	William W. Lambert	George G. Thompson
Urie Bronfenbrenner	Ronald Lippitt	Beatrice B. Whiting
Roger Brown	Icie G. Macy	John W. M. Whiting
Eleanor J. Gibson	Benjamin Mazalot	Herbert F. Wright
William E. Henry	Howard V. Meredith	Leon Yarrow
Lois Wladis Hoffman	Daniel R. Miller	Marion Radke Yarrow

1960 Approx. 1056 pages . $15.50*

**Textbook edition available for college adoption.*

Send for examination copies.

JOHN WILEY & SONS, Inc. 440 Park Avenue South, New York 16, N.Y.